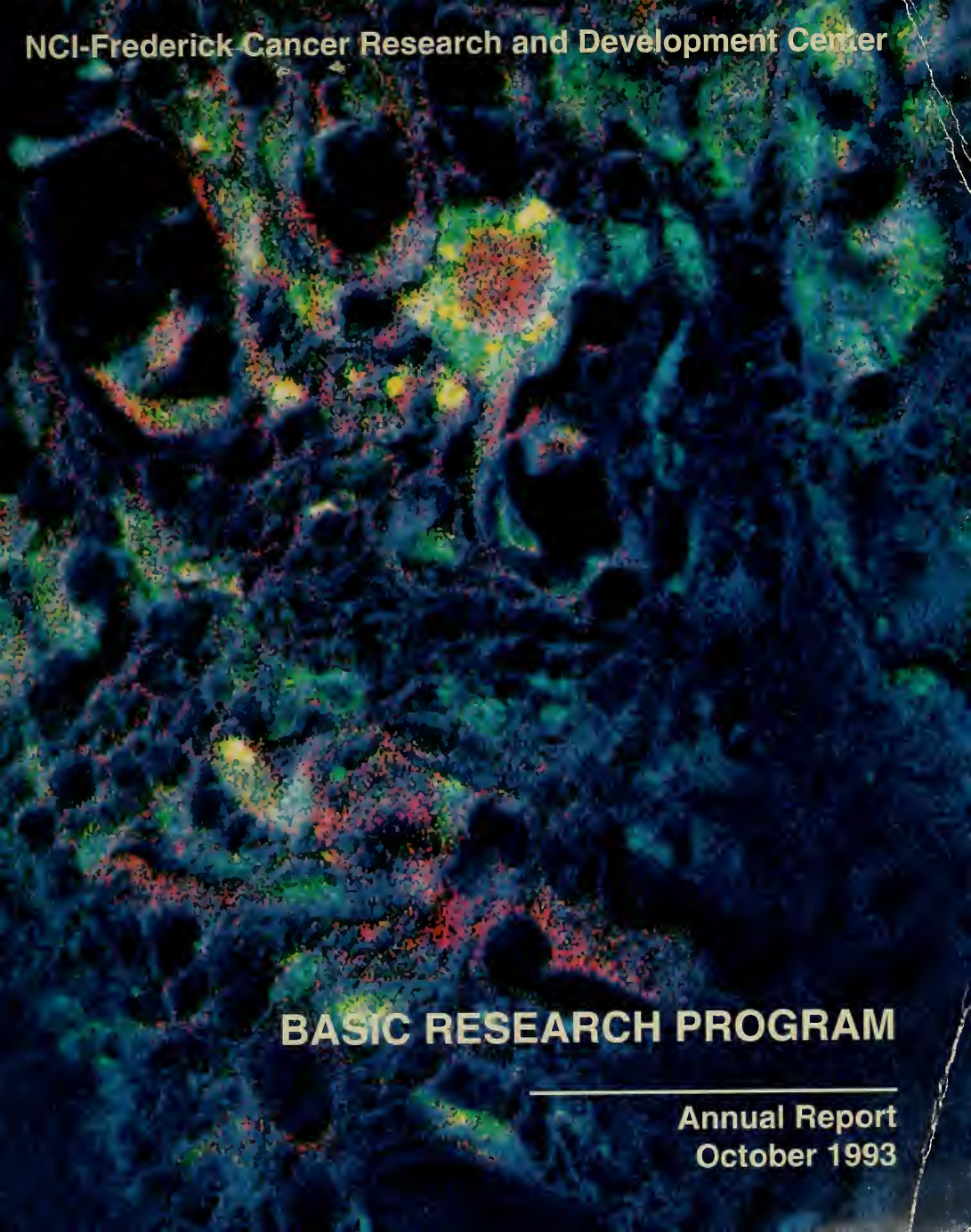
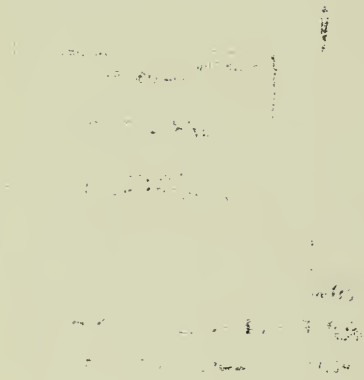
NCI-Frederick Cancer Research and Development Center



BASIC RESEARCH PROGRAM

Annual Report
October 1993



ON THE COVER:

A paraffin-sectioned human primary soft tissue sarcoma visualized by confocal laser scan microscopy (Molecular Oncology Section, Molecular Mechanisms of Carcinogenesis Laboratory). The front cover shows immunohistochemical staining for the met *proto-oncogene receptor (Met, green) and its ligand, hepatocyte growth factor/scatter factor (HGF/SF, light red), overlaid on the Nomarski image (blue) to reveal the cellular morphology. The areas of receptor/ligand colocalization are presented in yellow. The presence of discrete areas of green (Met), red (HGF/SF), and yellow (receptor plus ligand) staining indicates that both paracrine and autocrine signaling pathways can exist within the same tumor. The heterogeneity of Met and HGF/SF expression in the tumor may indicate tumor progression. For comparison, the back cover illustrates only the Nomarski image of the tumor section shown on the front cover.*

National Cancer Institute
Frederick Cancer Research and Development Center

BASIC RESEARCH PROGRAM

ANNUAL REPORT

October 1993

The Basic Research Program is operated for the National Cancer Institute by Advanced BioScience Laboratories, Inc., under Contract No. NO1-CO-74101.

P.O. Box B, Frederick, Maryland 21702-1201

Table of Contents

Introduction

This year marks my tenth anniversary with the Basic Research Program (BRP) and I take great pride in the many scientific accomplishments that members of our Program have made during this decade. Our major objective in fulfilling our mission for the National Cancer Institute has been to establish a program in basic research dedicated to scientific excellence. We have attracted world-class scientists who have developed independent programs that emphasize basic research in molecular biology, molecular genetics, and macromolecular structure. This expertise is essential for identifying, isolating, and characterizing genes responsible for human diseases, like AIDS and cancer; for understanding how these genes affect normal cellular processes; and for elucidating the molecular structures of the gene products that contribute to the disease process. In investigating these questions, our research complements and enhances the other programs at the National Cancer Institute's Frederick Cancer Research and Development Center (NCI-FCRDC) and the National Institutes of Health (NIH).

Some important achievements of BRP members during the past decade include the discovery of genes that are responsible for inherited genetic disorders, including familial cancers such as neurofibromatosis (NF-1). We also have identified genes that are the regulators of cellular responses to growth factors, including the Trk family of neurotrophin receptors (see the 1991 Annual Report cover) and the hepatocyte growth factor/scatter factor receptor (Met). Our work is elucidating the regulation of signaling molecules, including the Raf-1 kinase (Plate A of the pictorial section) and the TrkB receptor (Plate B). We have discovered genes that encode transcription factors, including those related to the prototype C/EBP leucine zipper protein (CRPs, Plate C) and a novel member of the basic-helix-loop-helix-zipper family (*microphthalmia* protein, Plate D). Through these studies, our Program has contributed to the extraordinary advances in understanding proto-oncogene signaling and transcriptional regulation.

During the past decade, the members of our Mammalian Genetics Laboratory (MGL) have been instrumental in the development of a comprehensive molecular linkage map of the mouse genome. They have placed over 1400 genes on this map, which is the paradigm for mammalian genome maps (see the October 1, 1993, issue of *Science*). Members of the Macromolecular Structure Laboratory (MSL) have made important contributions to determining the three-dimensional structures of biologically relevant molecules, including enzymes that are basic components of cellular metabolism (liver aldolase, Plate E) and those with proven efficacy in cancer therapy (asparaginase, Plate F). The structure of human immunodeficiency virus (HIV) protease was solved by members of the MSL (see Plate G and the 1989 Annual Report cover), and the structure of a complex of HIV reverse transcriptase and double-stranded DNA was solved in a collaboration between scientists in the Molecular Mechanisms of Carcinogenesis Laboratory (MMCL) and at Rutgers University (see the 1992 Annual Report cover).

In addition to important developments in basic research, our Program continues to pioneer new technologies that can lead directly to clinical applications. A compound, O^6-benzylguanine, was discovered in the Chemistry of Carcinogenesis Laboratory and is currently undergoing toxicological testing as a prelude to its use as an adjuvant in the treatment of human brain tumors. We are developing new strategies that will improve the screening of anti-HIV and anticancer drugs, and we are applying our molecular and structural expertise to the design of more efficient and rational drugs. An independent measure of our success and productivity is our publication record (see Table 1).

TABLE 1. ABL-BRP publications

Year	Peer-reviewed	Other	Total
1987	54	26	80
1988	83	22	105
1989	95	23	118
1990	103	33	136
1991	128	24	152
1992	134	15	149
1993*	65	7	72

*Through 8/31/93

One of our greatest achievements has been the development of a successful postdoctoral fellowship training program. This program has grown from less than 10 fellows in 1984 to 122 in 1993. The tenure of postdoctoral fellows averages 3 to 4 years and more than 100 fellows have completed their training in the BRP since 1984. Our postdoctoral fellows represent our most important investment in the future and our greatest single resource in the fight against cancer.

Several events of the past year deserve special mention. Early this year, two high-field nuclear magnetic resonance (NMR) spectrometers were installed for the Macromolecular NMR Section, which was established in 1992, and Andy Byrd and his staff began their NMR studies to determine the structure of proteins involved in signal transduction and transcriptional regulation. After five successful years, Ron Rubin (Molecular Pharmacology Group) left the MSL to establish a Crystallography Laboratory at Parke-Davis Pharmaceutical Company. We wish Ron great success in his new position.

We were fortunate to recruit James Resau from the Department of Pathology, University of Maryland School of Medicine, to head the Special Program in Confocal Laser Scan Microscopy (CLSM). Jim brings to our Program extensive experience in pathology and CLSM technology. He has already contributed significantly to a number of projects in the Program: most notably, he and members of the Molecular Oncology Section (MOS), MMCL, have examined the expression of the Met receptor and its ligand, HGF/SF, in human sarcomas (see the cover of this Annual Report) and cells undergoing a mesenchymal to epithelial cell conversion (see the MOS report, this Annual Report). We were also fortunate to recruit Marjorie Strobel, formerly a member of the MGL, as Scientific Assistant to the Director. She has already proven to be an invaluable asset to my office and to the entire Program.

The unexpected loss of Werner Kirsten, Associate Director of the NCI-FCRDC, saddened all of us. Werner was a close friend and a respected colleague, who contributed greatly to the development of the FCRDC. He is deeply missed.

George F. Vande Woude
October 1993

Editor's Note: During the past year, Dr. George F. Vande Woude was elected to the National Academy of Sciences. The members of the Program congratulate him on this special honor.

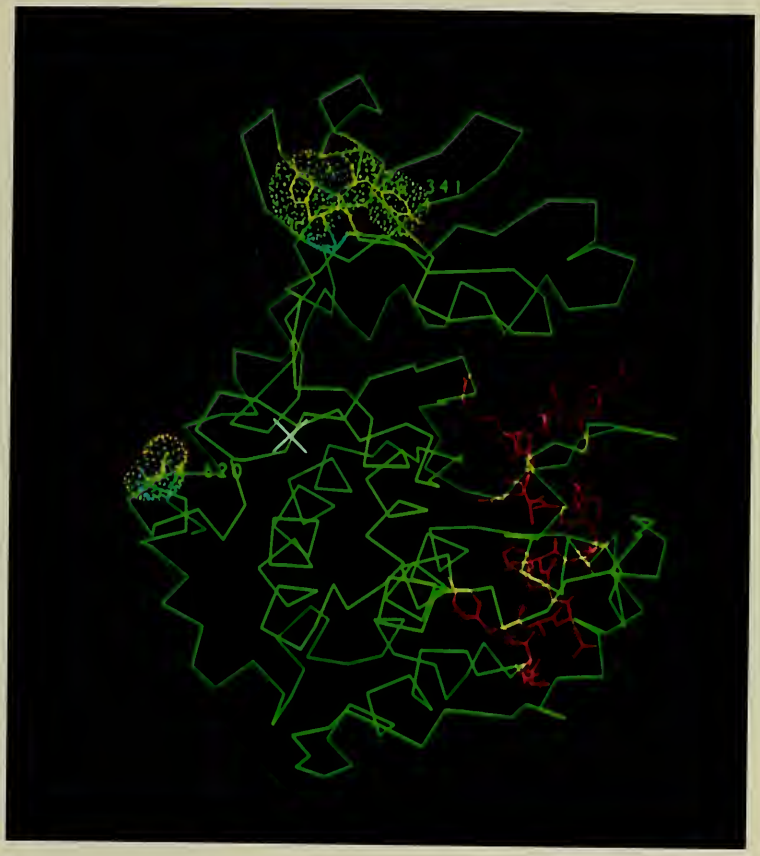


Plate A: Model of the catalytic domain of the Raf-1 kinase (green) bound to a peptide substrate (pink). Phosphorylation sites located within the catalytic domain (tyrosine 340 and 341 and serine 621) which were found to regulate the enzymatic and biological activity of Raf-1 are indicated by their van der Waals surfaces (colored spheres). Mutation of serine 621 to alanine inactivates the kinase, while mutation of tyrosine 340 or 341 to an acidic residue activates the kinase and results in a protein that can induce tumorigenesis. This study was carried out by Jack Fabian and Deborah Morrison, Cellular Growth Mechanisms Group. The Raf-1 structure was modeled by Ronald Rubin, Molecular Pharmacology Group.

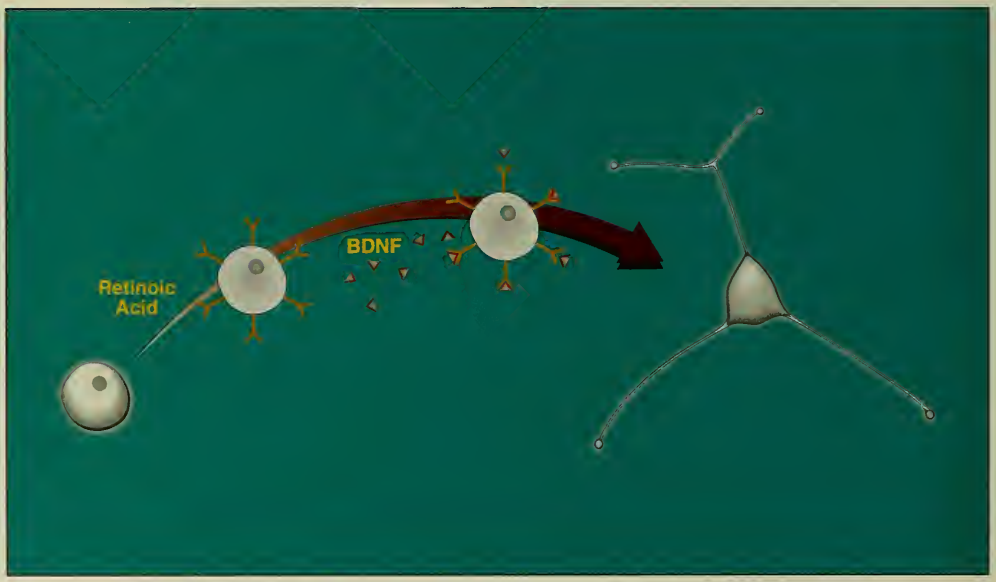
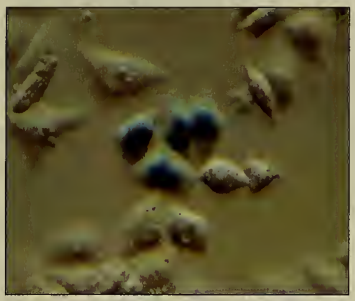
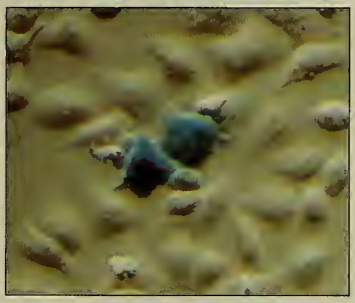


Plate B: Mechanism of neuronal differentiation of neuroblastoma cells. Human neuroblastoma is a pediatric tumor that in later stages is malignant and nonresponsive to chemotherapy. Cells derived from neuroblastoma tumors may be induced to stop growing and to differentiate into neuronal cells by treatment with retinoic acid and brain-derived neurotrophic factor (BDNF). Retinoic acid induces the expression of the TrkB receptor tyrosine kinase. Treatment of the TrkB-expressing cells with the TrkB ligand, BDNF, stimulates the activity of TrkB and induces neuronal differentiation. This work was carried out by David R. Kaplan (Eukaryotic Signal Transduction Group) and Carol Thiele (National Cancer Institute).





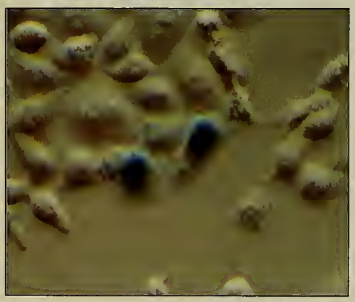


Plate C: The basic region of CRP1, a member of the C/EBP family of bZIP transcription factors, functions as a nuclear targeting signal (NTS) in addition to its previously established role as a DNA-contacting domain. The lacZ *gene of* Escherichia coli *was fused to the SV40 T-antigen NTS (top panel), the CRP1 basic region (bottom panel), or was left unchanged (center panel). The chimeric genes were then transfected into HeLa cells and the subcellular locations of the proteins visualized by histochemical staining. The basic regions of the other members of the C/EBP family also direct nuclear localization (data not shown). This work was carried out by Simon Williams and Peter Johnson of the Eukaryotic Transcriptional Regulation Group, as part of their ongoing studies of functional domains in the C/EBP-related proteins.*

Plate D: Mice with mutations at the microphthalmia *(*mi*) locus can exhibit a range of phenotypes, including pigment loss, diminished eye size, defective secondary bone resorption, reduced mast cell number, and early onset deafness. Homozygotes for the red-eye white mutation (*mi^{rw}*, top) have small red eyes and some pigmentation around the neck, while heterozygotes for the semidominant* microphthalmia white *(*Mi^{wh}*, middle) exhibit only a characteristic belly spot. Animals homozygous or heterozygous for the* microphthalmia-spotted *allele (*mi^{sp}*, bottom left) are virtually indistinguishable in color from wild-type mice, but* Mi^{wh}/mi^{sp} *mice (bottom right) are light yellow with dorsal and ventral white spots and pigmented eyes. In collaboration with Colin Hodgkinson and Heinz Arnheiter (National Institute of Neurological Disorders and Stroke), members of the Mammalian Genetics Laboratory showed that these diverse phenotypes and allelic interactions are caused by mutation of a novel member of the basic-helix-loop-helix-zipper (bHLH-Zip) family of transcription factors. (Photography courtesy of Karen Moore)*

Plate E: Ribbon diagram of rabbit liver aldolase. This enzyme is a tetramer and each subunit in the figure is represented by a different color. The X-ray structure has been refined to an R-factor of 21% with 2.1 Å data. This work is a collaboration between B.K. Sathyanarayana (Macromolecular Structure Laboratory) and the laboratory of Jurgen Sygusch (University of Montreal). They are also conducting similar refinement studies on rabbit muscle aldolase.

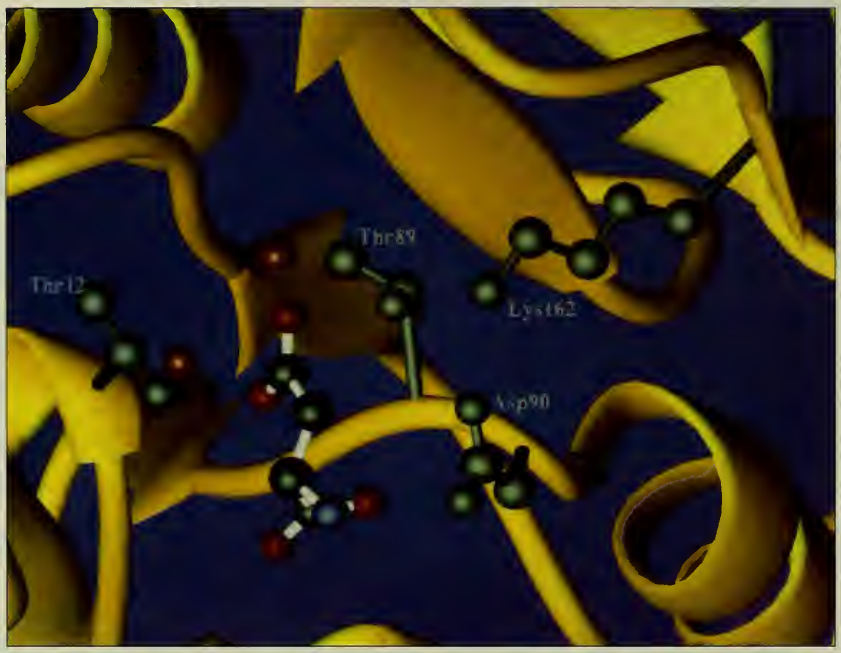


Plate F: A close-up of the active site of Escherichia coli *asparaginase, an enzyme used for cancer therapy. The product of the asparaginase reaction, aspartate, is shown surrounded by residues (labeled) and water molecules (individual red spheres) that are important for catalysis. The analysis of the active site of* E. coli *asparaginase was performed by members of the Enzyme Structure Section.*

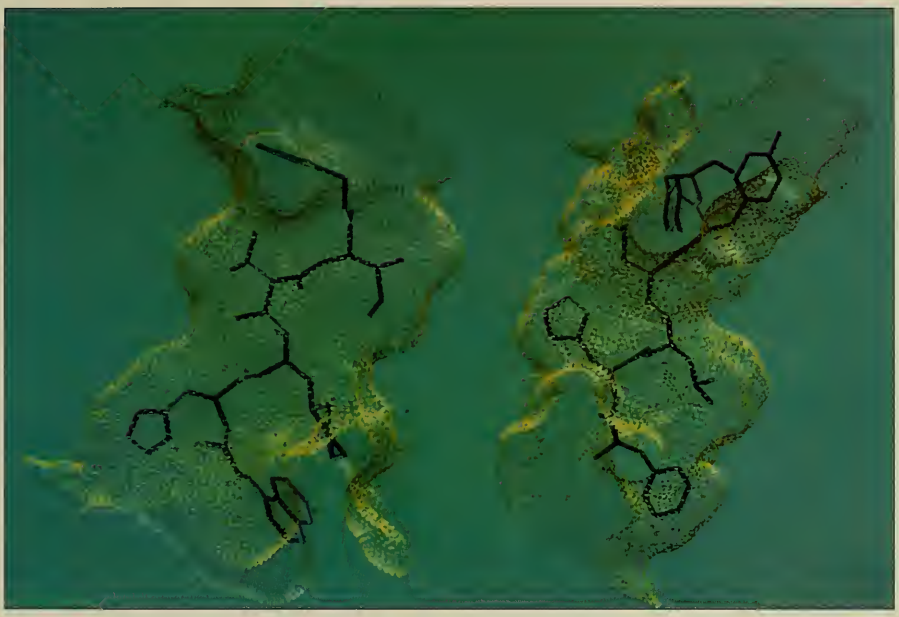


Plate G: The solvent-accessible surfaces (orange dots) of the substrate-binding clefts of the homologous enzymes HIV-1 protease (solved by members of the Macromolecular Structure Laboratory) and human renin (solved at Birkbeck College, London), with inhibitor coordinates superimposed. The surfaces were calculated on the basis of the structures of these enzyme-inhibitor complexes that were solved by X-ray crystallography. These studies were performed by members of the Enzyme Structure Section.

Laboratory Reports

Macromolecular Structure Laboratory

Alexander Wlodawer

Christopher Michejda
R. Andrew Byrd
Ronald Rubin

The mission of the Macromolecular Structure Laboratory is the analysis of atomic-level structures of proteins, nucleic acids, and small molecules of biological importance, as well as the design and synthesis of biologically active molecules and drug precursors. The techniques used by the laboratory include X-ray crystallography and nuclear magnetic resonance spectroscopy (NMR), as well as many approaches widely utilized in the fields of chemistry and molecular biology. The newest section of the laboratory, Macromolecular NMR, is now fully operational, and its two instruments (600 MHz and 500 MHz spectrometers) have been installed. Thus, the laboratory is pursuing its focus on broad utilization of physical techniques for the determination of macromolecular structures and for using these structures in rational drug design.

The full reports from the laboratory are provided by the Molecular Aspects of Drug Design Section, headed by Christopher Michejda, the Molecular Pharmacology Group, headed by Ronald Rubin, and the Enzyme Structure Section, headed by Alexander Wlodawer. A shorter report outlining the preliminary results and future plans of the Macromolecular NMR Section, headed by Andrew Byrd, is also provided.

The Molecular Pharmacology Group has concentrated on studies of interferons from several sources, on crystallographic investigations of a transforming growth factor, and on preliminary crystallographic experiments on a number of other systems. The Enzyme Structure Section has continued the process of solution and refinement of the crystal structures of the anticancer drugs, L-asparaginases, and has completed five structures at different degrees of detail. This section has also continued its investigation of the structures of retroviral proteases and of two cytokines, interleukin-4 (IL-4) and monocyte chemotactic factor. Furthermore, this section has conducted a broad survey of the common features of four-helix cytokines similar to IL-4 and of their mode of receptor binding. The Molecular Aspects of Drug Design Section has continued its work on the development of new chemotherapeutic agents based on the triazene moiety and has continued studies of the gastrin receptor. Finally, collaborative projects between Dr. Rubin's group and Dr. Wlodawer's section, involving crystallization studies of genetically truncated retroviral integrases as well as analysis of inhibitor binding to *Erwinia* asparaginase, have been continued.

Enzyme Structure Section

Alexander Wlodawer

Jerry Alexandratos
Alla Gustchina
Mariusz Jaskólski
Jacek Lubkowski
Dwayne McCully
Alexander Pavlovsky
J.K. Mohana Rao
Amy L. Swain
Narmada Thanki

The main objective of the Enzyme Structure Section is to investigate the three-dimensional structures of enzymes and cytokines and to correlate structure with function. This year, the major accomplishment was solving the structure of human recombinant interleukin-4 (IL-4) and correlating it with the structures of other helical cytokines. We have now solved five structures of L-asparaginases [from *Escherichia coli* (EcA), *Erwinia chrysanthemi* (ErA), *Acinetobacter glutaminasificans* (AGA), *Pseudomonas* 7A (PGA), and *Wolinella (Vibrio) succinogenes* (WA)], yielding insights into the plasticity of the active sites of these enzymes. We have continued our investigations of the structure and specificity of protease (PR) from Rous sarcoma virus (RSV) and human immunodeficiency virus type 1 (HIV-1). In these projects, the main focus has shifted from structure determination to interpretation; nevertheless, new experimental data have also been obtained. Molecular replacement studies of a human cytokine related to interleukin-8, namely monocyte chemotactic factor (MCAF), have been continued. Furthermore, we have analyzed crystal structures of several complexes of bovine pancreatic ribonuclease (RNase A) with dinucleotides to shed light on the mechanism of enzymatic activity of this important enzyme.

Complexes of RNase A with Dinucleotide Inhibitors

RNase A has served as a paradigm of enzymes for over 60 years. Although its structure has been solved by a number of laboratories and in a variety of crystal forms, the enzymatic mechanism is not yet completely understood. To probe the details of this mechanism, we decided to cocrystallize RNase A with a number of substrate-based inhibitors consisting of modified dinucleotides. Two of the inhibitors were provided by Marat Karpeisky (Engelhardt Institute of Molecular Biology, Moscow, Russia). Both of them included a noncleavable phosphorane group placed between two nucleotides, mimicking a good substrate (UpcA) and a poor substrate (UpcU). The third inhibitor, dCpdG, was obtained commercially, and it represents a derivative of the least effective substrate in this series.

Our cocrystals of RNase A with these inhibitors were isomorphous with the classic monoclinic form, which was used to refine the structure of the apoenzyme at high resolution. We collected data to 1.4 Å resolution for the complex with UpcA, 1.5 Å for UpcU, and 1.6 Å for dCpdG, and we refined the structures of these complexes to the R-factors of 16.8%, 15.1%, and 14.8%, respectively. The only inhibitor that was found to bind in a predicted manner was UpcA. Although very clear electron density corresponding to both bases was observed in the B1 and B2 pockets, the linkage between the nucleotides was not well ordered, and the positions of the atoms in that part of the structure were not clearly defined. Only the B1 pocket was clearly occupied in the complex with UpcU, while the B2 pocket was occupied only by several water molecules. This must indicate that the second half of the molecule interacts with the enzyme only for a small fraction of the time, and thus can explain the relatively poor catalytic efficiency of RNase A for a pyrimidine-pyrimidine substrate. As shown previously for dCpdG soaked into the crystals of the RNase A apoenzyme, this inhibitor binds in a completely

different mode, which does not correspond to productive binding of the substrate. The inorganic phosphate in the active site is not replaced, and it is the guanine that binds in the B1 site, although in a way very distinct from the mode of binding of a uridine in the other two compounds. The significance of this mode of binding is unclear, but it points out the danger of postulating that all related inhibitors bind similarly. Such postulates, however, are often invoked as the basis of rational drug design.

Crystal Structures of Amidohydrolases

Structural and Enzymatic Studies of Asparaginases. Amidohydrolases catalyze the hydrolysis of asparagine and/or glutamine to aspartate/glutamate. These enzymes are commonly present in lower organisms and are expressed primarily when the organism is utilizing amino acids for its main energy source. To date, we have solved the crystal structures of five bacterial amidohydrolases, two of them in the last year (see Table 1 and the next section), which exhibit a wide range of specificity for the two natural substrates. Amidohydrolases that exhibit a substrate preference for asparagine (asparaginases) are used for treatment of acute lymphoblastic leukemia. When administered in combination with other drugs, such protocols can bring the patient into remission. Unfortunately, adverse side effects are associated with this form of therapy. One of the problems is that the enzymes currently used in therapy (EcA and ErA) are not completely specific for asparagine. Depletion of circulating glutamine levels due to asparaginase administration may be a cause of toxic side effects.

Bacterial amidohydrolases of this family are active only as tetramers of chemically identical, two-domain subunits. Each of the four active sites of the tetramer consists of loops contributed by the N-terminal domain of one subunit and the C-terminal domain of an adjacent subunit. Although the enzymes discussed here are isolated from different sources, the members of this family are structurally so similar that they can be considered and compared as a group of extensive mutants. Of the five amidohydrolase structures that have been determined, the EcA and ErA structures are the highest quality, with regard to resolution of data and clarity of electron density maps. Both of these enzymes were found to have ligands bound in their active sites. The structures of AGA, PGA, and WA have been refined to the present limits of X-ray data, with the sequence for the last two being incomplete and unknown, respectively. Because of these limitations, and probably due to the lack of ligands in the active sites, the electron density in these regions of the AGA and PGA maps is sparse. The active site of AGA with glutamate bound was modeled with reasonable confidence based on the structures of EcA and ErA. The most we can say about WA is that the molecular replacement solution is correct and that the structure will

TABLE 1. Amidohydrolase refinement statistics

	E. coli EcA	*Acinetobacter* AGA	*Erwinia* ErA		*Pseudomonas* PGA	*Wolinella* WA
Resolution (Å)	2.3	2.9	1.8	1.8	2.0	3.0
Number of reflections	56390	7403	99552	81855	82323	12636
Asymmetric unit content	tetramer	monomer	tetramer	tetramer	tetramer	dimer
Ligand bound	4 aspartates	none	4 sulfates	4 aspartates	none	none
Number of atoms						
Protein	9708	2497	9860	9860	10008	4854
Water	601	0	1070	797	0	0
Other	36	0	20	36	0	0
R-factor	0.145	0.171	0.155	0.164	0.219	0.236
rms deviations from ideality						
Bond (Å)	0.014	0.020	0.015	0.015	0.019	0.014
Angle distances (Å)	0.043	0.050	0.050	0.050	0.065	0.043

be easily refinable when the sequence is determined. The comparisons in the study thus far are based on this modified model of AGA and the X-ray structures of EcA and ErA.

The residues that are involved in enzymatic activity have been identified by site-directed mutagenesis studies (conducted in the laboratory of Klaus Röhm, Phillips University, Marburg, Germany) in combination with crystallographic structural studies (Figure 1). Superposition of the ErA and the modeled AGA tetramers on EcA show that the amidohydrolase structures are very similar. However, there are some striking differences in the nature and orientation of some active-site residues. Of the several residues that line the surface of the active-site cavity, there are four that are different either in identity or orientation among the asparaginases (labeled with parentheses in Figure 1), and these differences affect the constellation of the active site and ultimately the activity of the enzyme. (In the modification of the AGA model, the positions of the four nonidentical residues in the active site were not significantly altered.) Although the nature of the residues in the active sites of ErA and AGA is identical, the orientation of Glu283 is different in ErA because of a 5-residue deletion.

The kinetic parameters for the amidohydrolases are given in Table 2. AGA and PGA bind both asparagine and glutamine substrates more tightly than do EcA and ErA (based on the K_M values) but are less discriminating. EcA and ErA both bind asparagine considerably more tightly than glutamine. WA, the most specific amidohydrolase for asparagine, was found to hydrolyze glutamine at a rate of only 0.015% of that for asparagine.

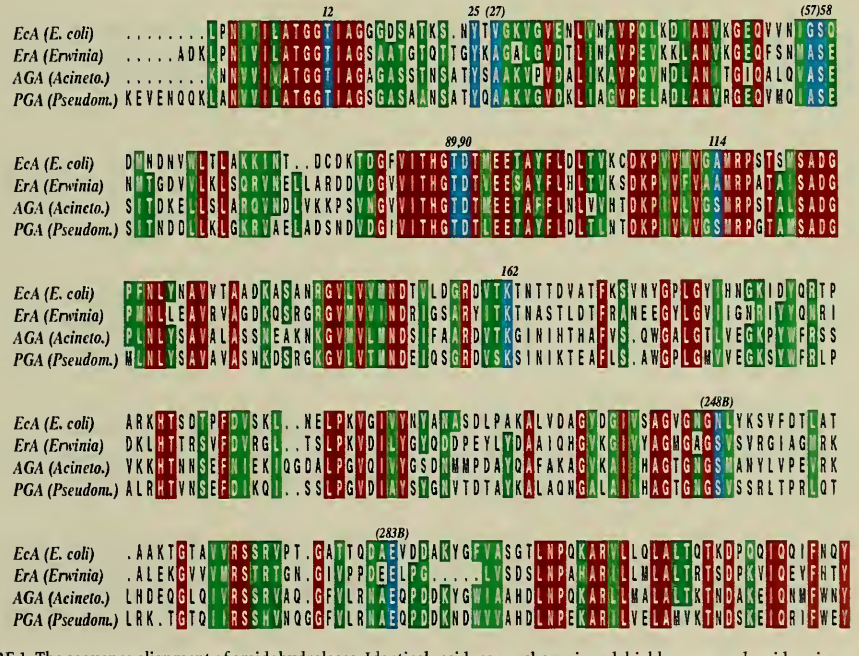


FIGURE 1. The sequence alignment of amidohydrolases. Identical residues are shown in red, highly conserved residues in green, and three identical residues out of four in dark green. The active-site residues are shown in blue and are numbered according to the sequence of EcA. The nonidentical active-site residues, with respect to residue type or side-chain position, are in parentheses. (Although residue 114 is not identical, it is not designated with parentheses because only its carbonyl oxygen contributes to the active site.)

TABLE 2. Kinetic parameters for amidohydrolases

Enzyme	Substrate	K_M (μM)
EcA	Asn	11.5
	Gln	6250
ErA	Asn	15
	Gln	1400
AGA	Asn	4.8
	Gln	5.8
PGA	Asn	4.4
	Gln	4.6
WA	Asn	47.8
	Gln	NM†

*ND, not determined.
†NM, not measurable.

While the sequences and the fold of this family of enzymes are highly conserved, the structural diversities at the active sites are reflected in the various binding affinities, rates, and specificities of the enzymes. The differences in residues Val27(Ala) and Gly57(Ala) appear to be unimportant, since the total volume occupied by these spatially adjacent side chains is similar and the differences compensate for one another. We therefore expect the differences in activities of the amidohydrolases to be manifested by Glu283 and Asn248(Ser). From the comparison of the K_M values, it appears that the position of Glu283 is not important for the binding of substrate, but it may be important for the rate of hydrolysis. In the EcA active site, there are close contacts among the amine group of the bound aspartate, the carboxyl group of Glu283, and the amide group of Asn248. Perhaps these residues are important for proper positioning of the ligand for efficient hydrolysis. While the Ser(Asn248) side chain of ErA, AGA, and PGA is capable of hydrogen-bonding to the amine group of the bound ligand, Asn248 of EcA has a better reach. Our conclusions await confirmation derived from ongoing X-ray analysis of glutaminases with ligands bound in their active sites and of the most specific asparaginase, WA.

Amidohydrolase Structures in Progress: PGA and WA. This year, we solved the structures of PGA and WA, two amidohydrolases with very different specificities for amide substrates. PGA exhibits higher affinity for glutamine than for asparagine, whereas the binding constant of WA for glutamine is too low to be measured. The structures are expected to reflect the contrasting activities of these two otherwise very similar enzymes.

PGA crystallizes in space group $P2_12_12_1$ with the unit cell parameters $a = 118.26$ Å, $b = 130.75$ Å, $c = 85.09$ Å. The asymmetric part of the unit cell consists of four chemically identical protein subunits. These characteristics and noncrystallographic symmetry information were reported several years ago based on 2.4 Å resolution X-ray data. This data set was collected with three different devices (a diffractometer and two area detectors) and is called here the "old" X-ray data set. The identities of only 26 amino acids at the N terminus of PGA were previously reported in the literature. This information showed that the protein chain of PGA is significantly longer at the N terminus compared to the other known amidohydrolases. The crystal structure of this enzyme was solved by molecular replacement using the "old" X-ray data set and the crystal structure of EcA as a model. Rigid body, conventional, and simulated annealing refinements were performed using the program X-Plor followed by several cycles of conventional refinement with PROFFT. These procedures decreased the R-factor from an initial value of 0.34 to 0.29.

Since our knowledge of the amino acid sequence was incomplete, during the subsequent steps the residue identities were estimated from electron density maps. Information about the sequences of other asparaginases, especially for the highly conserved regions of these enzymes, was also useful in making these assignments. Sequence corrections, manual geometry improvements, refinement processes, and electron density map recalculations gradually improved the model. At this stage, the correspondence between electron density maps and the model was reasonable for ~75% of the protein chain. The improvement was reflected in the R-factor of 0.19. Unfortunately, in several critical regions, including the active site, the electron density maps were unclear and further improvement was difficult without more information about the amino acid sequence. The current results of our PGA study suggest that the active sites are not occupied by any ligand, similarly to what was observed for AGA.

Although the complete amino acid sequence is not yet available, it is very close to being completed by Terry Copeland (Special Program in Protein Chemistry).

Only 27 of 336 residues (assuming no further deletions and insertions) remain unidentified. The sequence of PGA is most similar to that of AGA. Of 309 residues already identified for PGA, 187 are identical to the analogous residues of AGA, while for EcA and ErA the numbers are 142 and 135, respectively. All positions that have been determined through biochemical studies as being crucial for activity are occupied by identical amino acids in PGA. Six positions in the sequence, which were identical for other amidohydrolases (Figure 1), are different in the case of PGA (38, 74, 126, 161, 272, and 311 according to EcA numbering). In three of these substitutions, the differences were quite significant: Asn74Ala, Pro126Met, and Arg272His. An error in the identity of one amino acid in the N-terminal fragment of the PGA sequence published previously has been identified.

While these investigations were in progress, we received a sample of PGA from Herman Ammon (University of Maryland, College Park). This sample enabled us to determine the amino acid sequence as described above and to grow new crystals in order to obtain a more consistent and complete X-ray data set. Crystals of PGA that were grown using the published protocol were not sufficiently stable to collect X-ray data. To stabilize the crystals, we soaked them in 0.03% to 0.04% glutaraldehyde in the solution from which the crystals were grown (40% MPD and phosphate buffer). We collected a data set at 2.0 Å resolution, using a rotating anode X-ray generator with an R-axis II detector. Currently, the reduction of experimental data is in progress and we believe that the "new" X-ray intensity data with complete amino acid sequence information for PGA and the current model will make it possible for us to obtain a better picture of PGA, including the details of the active-site regions.

The structure of WA is in progress, with only a refined molecular replacement solution available at this time. We have positioned the model of an EcA dimer in the WA asymmetric unit and refined the structure to an R-factor of 0.24. The resolution of the WA crystal data is only 3.0 Å and the sequence is completely unknown. We are collaborating with Professor Röhm in cloning and expressing the gene that encodes WA, in order to determine its amino acid sequence. This collaboration should also provide us with protein for growing crystals necessary to collect diffraction data at higher resolution.

Investigations of ErA. The following investigations were carried out in collaboration with Maria Miller (Molecular Pharmacology Group) and Michael Gribskov (San Diego Supercomputing Center, San Diego, CA). ErA is a homotetramer of 327-residue subunits with approximate 222 symmetry. Each ErA subunit consists of two domains connected by an extended linker (residues 197 to 218). The larger N-terminal domain comprises two subdomains. The first 156 residues form a parallel α/β structure similar to that of dehydrogenases, flavodoxin, and the two domains of arabinose-binding protein, with a strand connectivity of [-1x,+2x,+1x,+1x]. Residues 157 to 196 of the N-terminal domain form a two-layered, antiparallel β-sheet. Three of the five strands in this subdomain continue the β-sheet from the N-terminal subdomain, giving rise to an eight-stranded, mixed β-sheet with the the first five strands being parallel and the last four being antiparallel. The smaller C-terminal domain consists of a four-stranded, parallel β-sheet with [+1x,+1x,+1x] connectivity. This parallel sheet, as in the N-terminal case, has two α-helices on either side.

In L-asparaginases, the crossover between Nβ4 and Nβ5 in the N-terminal domain is left-handed. Such crossovers have been observed only rarely in protein structures (for example, in subtilisin, acetylcholine esterase, and steroid dehydrogenase). In subtilisin, residues in the left-handed crossover are essential for the function and activity of the enzyme. This also appears to be the case in the Type II L-asparaginases in which the left-handed crossover forms part of the active site (see below). Furthermore, the amino acid residues that comprise the left-handed crossover are evolutionarily conserved (Figure 1), probably because residues 128 to 132 provide a substantial part of the interface between the pairs of dimers in the tetramer. These residues are in contact with their symmetry-related cognates from the opposite dimer (i.e., they are related by a noncrystallographic dyad axis).

The difference Fourier map computed with the data collected on crosslinked ErA crystals (soaked in aspartic acid solution buffered at pH 5.0) revealed the location and orientation of the bound aspartate, and the subsequent refinement proved that the substrate was present in the active site at high occupancy.

As shown in Figure 2, the side-chain atoms of the bound aspartic acid are located between the hydroxyl groups of

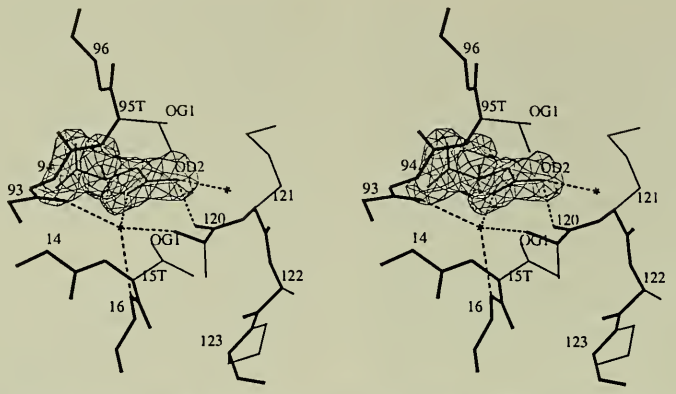


FIGURE 2. Stereoview of ErA with bound aspartic acid in the initial (F_o-F_c) electron density contoured at the 3σ level. Selected hydrogen bonds between the reactive end of the substrate and its environment are shown as dashed lines, two water molecules are indicated by asterisks, and the enzyme main chain is indicated by thick lines. For clarity, not all side chains (thin lines) are shown.

Thr15 [(12_{EcA}); to simplify comparisons between ErA and EcA, the EcA residue numbers are given in parentheses] and Thr95(89_{EcA}) and near the main-chain atoms of Ala120(114_{EcA}), the first residue of the left-handed connection between Nβ4 and Nβ5. The catalytic portion of the active site is therefore made up of the three loops linking Nβ1 to Nα1 (residues 14 to 35), Nβ3 to Nα3 (residues 93 to 97), and Nβ4 to Nα4 (residues 119 to 130). The β-carboxyl oxygens of the substrate are almost coplanar with the carbonyl of Ala120(114_{EcA}), which can act as an acceptor in a hydrogen bond with the aspartate $O_{\delta 2}$, suggesting that the bound aspartate is protonated. At neutral pH, the apposition of the charged aspartate carboxylate group and the main-chain carbonyl oxygen of Ala120(114_{EcA}) generates a strong repulsion that helps to release the product of the enzymatic hydrolysis. Two water molecules are also located in the same plane, forming a hydrogen bond network aiding in the proper positioning of the substrate. One of the water molecules anchors the β-carboxylate oxygen of the substrate to the main-chain atoms of the enzyme by hydrogen bonds: as a proton donor to $O_{\delta 1}$ of the substrate and carbonyl oxygen of His93(87_{EcA}) and as an acceptor from main-chain NH groups of Ile16(13_{EcA}) and Ala120(114_{EcA}). $O_{\delta 1}$ of the substrate side chain is sandwiched between the main-chain nitrogen of residues Thr15(12_{EcA}) and Thr95(89_{EcA}). This is similar to the "oxyanion hole" seen in serine proteases and presumably acts to stabilize the partial negative charge that develops in the tetrahedral intermediate of the formation of the acyl-enzyme complex. When asparagine is a substrate, the carbonyl oxygen presumably occupies the position of $O_{\delta 1}$, while the amide nitrogen occupies the position of $O_{\delta 2}$.

The α-carboxylate end of the substrate is held in position by a network of hydrogen bonds and charge interactions. The carboxylate group is oriented by hydrogen bonds to the main-chain nitrogens of Ser62(58_{EcA}) and Asp96(90_{EcA}) and to the side-chain Oγ of Ser62(58_{EcA}). The α-amino group interacts with the acidic residues Glu63(59_{EcA}) and Asp96 (90_{EcA}). This interaction probably accounts for the observation that the substrate α-amino group must be protonated.

The amino acid residues that form the surface of the active site are shown in Figure 3. Thr95(89_{EcA}), Lys168 (162_{EcA}), and Asp96(90_{EcA}) are connected by strong hydrogen bonds. Asp96(90_{EcA}) also forms a hydrogen bond with Ser254(248_{EcA}) from another subunit. As a result of this string of interactions, this side of the active site is rigid. In contrast, both Thr15(12_{EcA}) and Tyr29(25_{EcA}) moved slightly from the positions observed in the native structure, indicating that this side of the active-site cleft is somewhat flexible. In addition to its close proximity to the substrate, the hydroxyl of Thr15(12_{EcA}) might also act as a donor in a hydrogen bond with Tyr29(25_{EcA}).

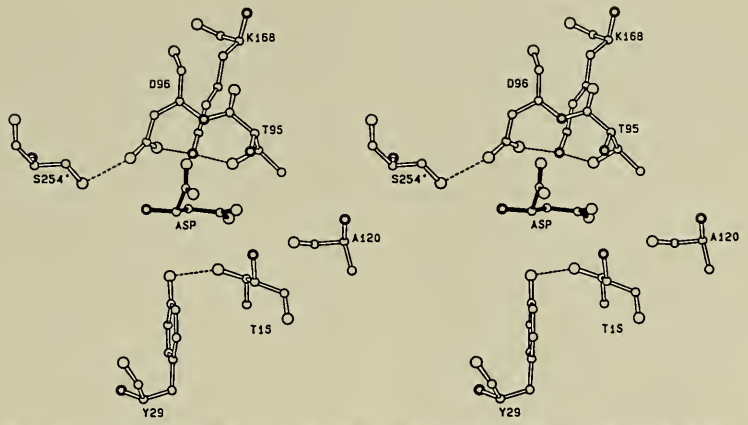


FIGURE 3. Constellation of residues forming the ErA active-site surface. The bound aspartic acid is shaded with the β-carbonyl extending towards the right. Note the location of the γ of Thr15 and Thr95 below and above the Cγ of the substrate. No other side chains closely approach the C of the substrate.

Although none of the side chains of the left-handed crossover is part of the active-site surface, Met121(115_{EcA}) and Pro123(117_{EcA}) provide hydrophobic contacts for the methyl groups of Thr95(89_{EcA}) and Thr15(12_{EcA}), thus preventing rotation.

Mutation of either Thr95(89_{EcA}) or Thr15(12_{EcA}) diminishes enzymatic activity to about 0.01% of the wild-type activity. The positions of the amino acid residues in the active site suggest several possible pathways for the nucleophilic catalysis, but the nearly symmetric location of these threonine residues above and below the Cγ of the substrate suggests that one of them must be the catalytic nucleophile. This possibility is supported by the location of both threonines within two absolutely conserved regions of amino acid sequence located at residues 12(9_{EcA}) to 16(13_{EcA}) and 93(87_{EcA}) to 97(91_{EcA}) (Figure 1).

Comparison of Calcium-Dependent Conformational Changes in the N-Terminal SH2 Domains of p85 and GAP Defines a Distinct Property for SH2 Domains

The binding of platelet-derived growth factor (PDGF) and epidermal growth factor (EGF) to their respective cell surface receptors (PDGFR and EGFR) leads to receptor tyrosine kinase activation by a process involving receptor dimerization. Kinase activation is associated with receptor autophosphorylation and subsequent phosphorylation of intracellular substrates, some of which may be involved in the transduction of the mitogenic signal from the cell membrane to the nucleus. High-affinity association occurs between specific autophosphorylation sites in the tyrosine kinase receptors and specific conserved domains in the substrates. These domains are known as Src-homology region 2 (SH2), each consisting of around 100 amino acids. SH2 domains may be part of enzymes (e.g., PLC_γ, Src) or part of adaptors that couple to enzymes (e.g., p85, Sem-5/GRB2). The existence of SH2 domains has been established in many proteins by sequence alignment. A number of these signaling proteins, including phosphatidylinositol-3 kinase (PI-3K) binding subunit (p85), ras GTPase-activating protein (GAP), phospholipase C_γ (PLC_γ), and Src-like nonreceptor tyrosine kinases, have been shown to bind activated α- and β-PDGFRs.

The tyrosine kinase of PDGFRs contains a kinase insert (KI) domain. The tyrosine phosphorylation sites required for high-affinity binding of p85 and GAP have been identified within the KI domain of PDGFRs. The phosphotyrosine-containing pentapeptides that encompass the PDGFR binding sites for p85 and GAP have also been shown to compete specifically with activated PDGFR binding.

Moreover, circular dichroism and fluorescence spectroscopy methods have been used to show a significant conformational change in the p85 N-terminal SH2 (N-SH2) domain with the addition of a 17-residue polypeptide from the β-PDGFR KI domain phosphorylated on Tyr751.

The three-dimensional structures of v-Src, c-Abl, v-Lck, and p85 N-SH2 domains have been determined in several laboratories. It is evident that the SH2 domains of these distinct signaling molecules exhibit a similar tertiary fold, despite the high degree of primary structure variability, implying that they possess distinct physicochemical properties. The crystal structures of v-Src and Lck SH2 domains allowed a better structural alignment of the plethora of SH2 domain sequences. A perusal of this alignment strongly suggested calcium-binding motifs in the p85 N-SH2 domain but not in the GAP N-SH2 domain. We decided to investigate these properties experimentally.

The N-SH2 domains of PI-3K and GAP were expressed in M15 *E. coli* by using a six-histidine N-terminal tag in the pQE9 vector. These proteins were purified for structure-function studies by Ni-NTA chromatography and FPLC. We used a combination of empirical and theoretical methods to study physicochemical properties for those domains that may be involved in receptor signaling in vivo. The p85 N-SH2 domain stained dark purple and absorbed around 620 to 640 nm with "stains-all," a dye known to bind to calcium-binding proteins. This effect was not observed for the GAP N-SH2 domain. Circular dichroism analysis of the N-terminal SH2 domain of these proteins showed that p85 N-SH2, but not GAP N-SH2, undergoes a significant dose-dependent change in conformation in the presence of increasing calcium concentrations. Moreover, the conformational change of p85 N-SH2 induced by calcium could be replicated by addition of a phosphorylated hexapeptide (DYpMDMK) representing the αPDGFR binding site for p85. Limited proteolysis studies showed a significant calcium-dependent increase in protection of p85 N-SH2, but not GAP N-SH2, from degradation by subtilisin. Our results further indicated that holmium, a trivalent lanthanide ion that can substitute for calcium, also protected p85 N-SH2 from proteolysis, even at tenfold lower concentrations.

In vitro binding studies using purified preparations of activated αPDGFR showed that calcium did not affect the binding of GAP N-SH2 domains to activated αPDGFR. In striking contrast, we observed a marked increase in the binding of p85 N-SH2 domains to activated αPDGFR in the presence of calcium ions. Sequence comparisons and the molecular model of p85 N-SH2 domain based on the v-Src SH2 domain structure show a conserved arrangement of oxygen ligands contributing to two potential calcium binding sites within p85 N-SH2. These sites have been rationalized to surface loop regions (i.e., loops 2 and 4 and loops 6, 7, and 8), close to the N- and C-terminal α-helical regions, respectively. Together, our findings suggest that the conformation of p85 N-SH2, but not GAP N-SH2, is modulated by the presence of calcium ions. This modulation implies that calcium ions may regulate PI-3K (p85α) binding to αPDGFR in vivo and suggests that the N-SH2 domains of p85 and GAP may have distinct functions in receptor-mediated signal transduction.

A database scan using the p85 N-SH2 domain sequence as a probe showed a sequence similarity of 56% to rat cellular retinol binding protein (cRBP) over its entire length, including conservation of hydrophilic and hydrophobic amino acids. The three-dimensional structure of cRBP is not known. However, it belongs to a superfamily of lipid-binding proteins, one member of which is intestinal fatty acid-binding protein (I-FABP) and whose crystal structure is known. Despite significant primary and secondary structure similarity between p85 N-SH2 and cRBP, their topology is opposite, i.e., the central β-sheet in cRBP is perpendicular to the N- and C-terminal α-helices, whereas the central β-sheet lies parallel to the respective helices in the p85 N-SH2 structure. This observation suggests that proteins with very similar primary and secondary structures may fold differently to perform diverse functions, although both classes are ligand-binding proteins.

These results imply that despite their conserved three-dimensional structure, the SH2 domains found in a number of signaling molecules may have distinct functions in vivo. Cyclical changes in intracellular calcium concentration may play a critical role in the binding of p85 not only to αPDGFR but also to its catalytic subunit (p110) and insulin receptor substrate-1 in vivo.

Structural Studies of Helical Cytokines

Crystal Structure of IL-4. Many cytokines (proteins that influence cell growth and differentiation) have been identified so far. These molecules accomplish their biological

role by binding to specific receptors, which in turn generate intracellular signals, initiating developmental processes in the cells. In particular, the hierarchical assembly of hematopoietic cells is regulated by a number of cytokines, which bind to distinct, but related receptors. These hematopoietins include, among others, interleukins-2 through -7, granulocyte-macrophage colony-stimulating factor (GM-CSF), erythropoietin, and growth hormones. Medium- to high-resolution structures of several of the hematopoietins have now been determined by X-ray crystallography and NMR.

During the past year, we determined and refined the crystal structure of IL-4. Crystals of IL-4 belong to the space group $P4_12_12$ (the enantiomorph established on the basis of the anomalous scattering of the heavy atoms). Unit cell parameters are $a = b = 91.8$ Å, $c = 46.4$ Å, with one molecule present in the asymmetric unit. Derivative data were collected to the resolution of 2.7 to 3.5 Å (Table 3), while the native data extended to 2.23 Å and were reasonably complete [9169 out of 10,104 possible reflections measured, with 8223 reflections having $I > 1.5\,\sigma$ (1)].

An electron density map was computed with multiple isomorphous replacement (MIR) phases from five derivatives, three of which also provided usable anomalous data. The map was rather difficult to interpret, even though the four helices could be identified without ambiguity (Figure 4A). Attempts to improve the map by solvent flattening were not successful, despite the high solvent content of these crystals. The initial fit to this map was verified in a comparison with the NMR coordinates provided by Christopher Dobson, Oxford University, which also were useful in tracing some of the broken density belonging to the regions lacking regular secondary structure.

We subjected the initial atomic coordinates to three cycles of simulated annealing and to least squares refinement, resulting in the crystallographic R-factor of 0.218 for a model consisting of all 129 amino acids, as well as 33 water molecules (Figure 5). The quality of the final map is good (Figure 4B). The core of the molecule is formed by a left-handed helical bundle, consisting of helix A (residues 4 to 19), B (40 to 60), C (70 to 94), and D (109 to 126). The helices are connected by three loops forming the most flexible part of the molecule. Two β-strands that are part of the loops form a short fragment of antiparallel β-ribbon (residues 27 to 31 and 105 to 109) connected by four hydrogen bonds.

The core of the molecule between the helices is filled by residues with mostly hydrophobic character and is completed by several predominantly hydrophobic residues from

TABLE 3. Summary of crystallographic data collection and phasing for IL-4

Data set	Unique reflections (F>2σ)	Resolutions (Å)	Completeness (%)	R_{merge}*	R_{deriv}†	R_{center}‡	Phasing power§ iso (ano)	Heavy atom Number	Heavy atom Location
Native‖	8951	2.23	90.7	0.06					
Trimethyl lead acetate‖	4851	2.7	80.5	0.07	0.202	0.65	1.34 (1.48)	1	Asp31; Glu114
$K_2Pt(NO_2)_4$¶	2927	3.0	97.8	0.10	0.148	0.66	1.23 (1.34)	1	His58; His59
Iodine¶	3156	3.0	92.3	0.10	0.140	0.65	1.28 (1.43)	1	Tyr56
Gold sodium thiomalate¶	2067	3.2	67.7	0.11	0.150	0.68	0.98	1 2	Phe112 Thr18; Leu113
K_2PtCl_6¶	1934	3.5	87.4	0.10	0.178	0.71	0.95	1 2	Thr22; His76 Ser107; Arg115

The overall figure of merit is 0.524 for 5193 independent reflections with F>2σ(F) in resolution range 25 to 2.7 Å.

*$R_{merge} = \Sigma|1obs - <1>|/\Sigma<1>$

†$R_{deriv} = \Sigma|F_{PH} - F_P|/\Sigma F_P$

‡$R_{center} = \Sigma||F_{PH} \pm F_P|F_H|/\Sigma|F_{PH}-FP|$

§Phasing power for iso = $\Sigma F_H/E$ ano = $\Sigma 2F_H/E$

‖Data collected on R-AXIS

¶Data collected on Siemens

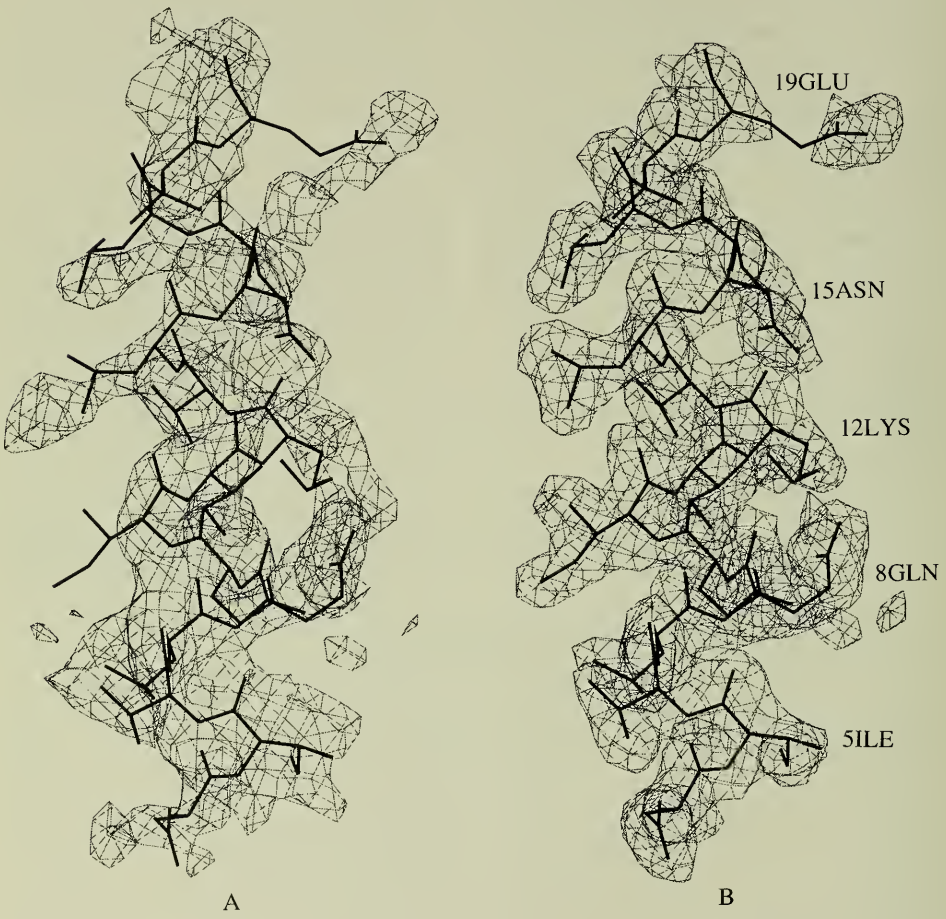


FIGURE 4. Electron density corresponding to helix A in IL-4 (residues 4 to 19) contoured at the 1 σ level, with the coordinates of the final model superimposed. (A) MIR map computed at 2.7 Å with the phases calculated as shown in Table 3. (B) $2F_o - F_c$ difference Fourier map at 2.25 Å resolution at the final stage of refinement.

FIGURE 5. Stereoview of the structure of IL-4. The main chain is presented by the C_α trace, while all side-chain atoms are shown in thin lines.

the irregular part of the structure. Some charged residues reside within the hydrophobic core, forming possible ion pairs. The side chain of Glu41 interacts with Lys123, while His59, Asp62, and Arg64 share an ionic interaction (the latter two residues are located outside of the helices).

A prominent feature of the structure is the string of positively charged residues (His74, Lys77, Arg81, Lys84, Arg85, Arg88). All of them are exposed and their charges are uncompensated by other residues, except for Lys84, which may form an ion pair with the side chain of Glu60. These residues, except for Arg88, are not involved in crystal contacts. Such formation of a positively charged "wall" may play a role in the interaction of IL-4 with its receptor.

Comparisons with Other Helical Cytokines. The structure of IL-4 revealed close similarity with other cytokines belonging to the four-helix structure family. Superimposed tracings of the C_α chains of IL-4, IL-2, and GM-CSF are shown in Figure 6. It is clear that the areas of best superposition include primarily the four-helix bundle. However, the similarity is not limited to helical regions alone, but extends to some connecting loops as well, in particular to the loop following helix A and another preceding helix D.

For slightly more than half of the structure, based on C_α atom positions, the deviations fall within about 1.5 Å rms from each other. When analyzed in greater detail, the superpositions are not equally close for all helices, with the best agreement found between helix D of each structure. It is also clear that while the positions of helices and the angles between them are quite well conserved, their lengths vary considerably from one structure to another.

To put the differences between these structures in another perspective, we compared the C_α coordinates of the X-ray and NMR models of IL-4. Differences between these coordinates could be due to a combination of inaccuracies of each structure and to genuine differences between the structures of proteins in crystals and in solution. For 126 out of 129 C_α atoms excluding the N terminus, the rms differences are 1.65 Å, while for the helical regions they range from 0.79 Å for helix C, 0.82 Å for helix A, and 0.97 Å for helix D to 1.78 Å for helix B. We may conclude that while some of the differences between the structures of the three cytokines might be due to the differences of crystal packing and to inaccuracies in structure determination, such influences are not major, since they are on the same order as the differences found in the same structure studied by

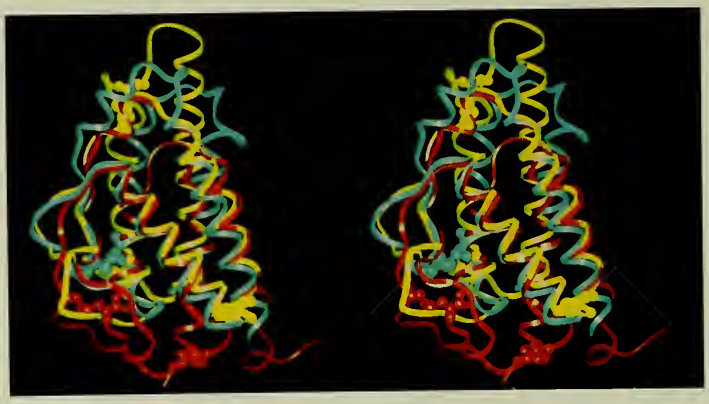


FIGURE 6. Stereoview of the superposition of the C_α traces of IL-4 (yellow), IL-2 (blue), and GM-CSF (red). All atoms in the side chains of cysteines forming disulfide bonds are shown as space-filling spheres.

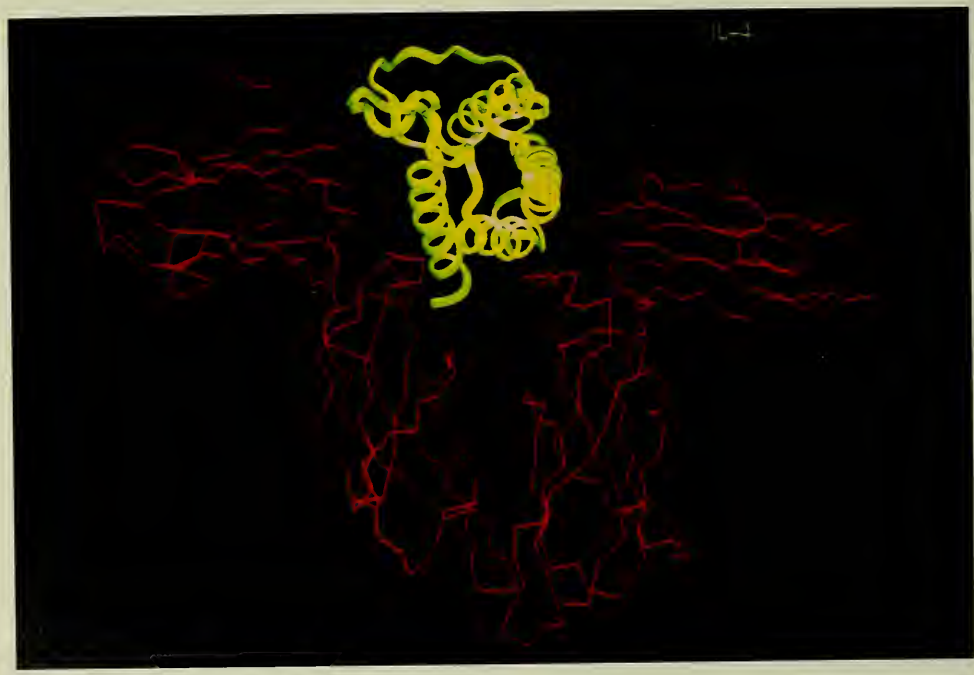


FIGURE 7. Modeled interactions of IL-4 with its receptor. The model of IL-4 (yellow) was superimposed on the C_α model of HGH (not shown), and the unmodified trace of HGHR is shown in red.

different means and in different environments. These differences are also much larger than those found in corresponding regions of the two crystallographically independent molecules of IL-2, where the rms differences are in the range of 0.2 to 0.3 Å.

Models for Cytokine-Receptor Interactions. The extracellular domains of cytokine receptors involved in the binding of hematopoietins share significant sequence and structural similarity. However, the only experimentally determined structure of an extracellular part of such a receptor that has been published is the result of studies of the complex of human growth hormone (HGH) with its receptor (HGHR). Each molecule of HGH is bound to two molecules of HGHR, with a large part of the surface area of the hormone involved in the contacts with its receptors. This area consists of two sites, site I (in contact with molecule 1 of the receptor) and site II (contacting molecule 2). Site I is the major binding site, with the interacting region 25% larger than for site II. In addition, considerable interactions between the two receptor molecules are also introduced by their dimerization.

The high degree of structural similarity between IL-4 (and other cytokines for which three-dimensional structures are known) and HGH make it possible to use the structure of the receptor in the complex of HGH/HGHR to identify those areas of the other cytokines that are likely to be involved in the interactions with their respective receptors. For this purpose, we superimposed the C_α atoms of IL-4 and HGH and found that this superposition does not lead to major steric clashes between IL-4 and the receptor molecules. We did not attempt to change the trace of the receptor to compensate for some insertions and deletions found when the sequences of HGH and IL-4 receptors were compared.

Figure 7 shows the model of IL-4 interacting with two molecules of HGHR. By mapping the residues in IL-4 that are in close proximity to the receptor in this simple superposition, we can suggest that site I includes residues 18, 19, and 21 in helix A; 31, 34, and 36 in the connecting loop AB; 105 to 108 in the connecting loop CD; and 110, 111, 114, 115, 117, 118, 121, 122, 125, 126, and 129 in helix D. For site II, the residues 6, 8, 9, 12, 15, and 16 in helix A and 74, 78, 81, 82, 85, and 88 in helix C could be involved in interactions with the extracellular domains of IL-4 receptors. Approximately 85% of the residues mentioned above belong to the regions with highly conserved three-dimensional structures in different cytokines, and some of these structures have similar or identical residues occupying structurally equivalent positions.

In the absence of experimental high-resolution crystal or NMR structures showing the interactions of different hematopoietins with their receptors, and with only limited mutagenesis data, the interpretation of the models based on similarity of primary through tertiary structures of the ligands has to be accepted as only hypothetical. In particular, even in the regions where considerable structural similarity is preserved, detailed interactions might differ due to compensatory changes on both the ligand and the receptor and to the possibility of accomplishing similar binding with residues that are not directly equivalent in their spatial location. This may indeed explain, for example, the apparent nonequivalency of Gln126 in IL-2 with Tyr124 in IL-4, despite some similarity of their biological roles.

However, attempts to map those regions of the cytokines that might be involved in the interactions with their receptors, as well as the regions on the receptors involved in cytokine binding, are quite reasonable. Similarities and differences discussed above may lead to designing experiments that could directly answer the questions regarding the nature of such interactions, and thus be of benefit in designing agonists and antagonists of these important signaling molecules.

Molecular Aspects of Drug Design Section

Christopher J. Michejda

Jeffrey Blumenstein
W. Marek Cholody
Grzegorz Czerwinski
David Farnsworth
Lidia Hernandez
Marilyn Kroeger Koepke
Carol Rouzer
Brigitte Schmidt
Ann Schmiedekamp
Richard H. Smith, Jr.
Nadezhda Tarasova

The objective of the Molecular Aspects of Drug Design Section is to use information obtained from biological, structural, and biochemical studies to design and synthesize new drugs against cancer and viral diseases. The synthetic effort is supplemented by studies on reaction mechanisms of candidate drugs. Computational chemistry is an important adjunct to our work and is indispensable in the design of drugs and in the a priori assessment of drug-macromolecule interactions. The chemical research carried out in this section is closely tied to the biochemical and biological research. Current work includes the characterization of the gastrin receptor (GR) in normal and neoplastic tissues, the isolation and structural and functional characterization of various human gastric proteases, development of a shuttle vector system to assess lethal DNA damage in vitro and in vivo, and assays for cytotoxicity in human tumor cell lines and in xenografted tumors. The design of experimental protocols is planned in joint sessions, so that the biologically oriented members of the section interact closely with the chemists, both in the planning stage and during the execution of the experiments. This report focuses on three topics in which this interaction becomes evident.

Gastrin Receptor-Mediated Drug Delivery

Last year, we reported our progress in the area of targeting drugs to specific receptors (1992 Annual Report). At that time, we had succeeded in the preparation of an acyltriazene linked to the C-terminal tetrapeptide that constitutes the recognition sequence of the peptide hormone gastrin. We found that the drugs CBS-4 and CBS-5, whose structures are shown in Figure 1, were able to bind to GR in freshly prepared guinea pig stomach membranes, whereas binding by CBS-2, the drug consisting of a dipeptide conjugated with the triazene, was much less effective.

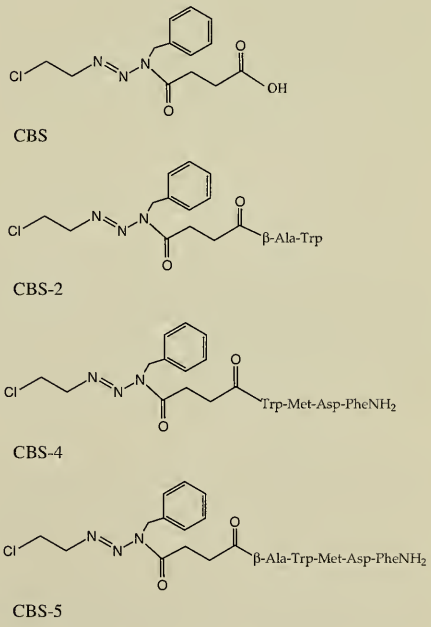


FIGURE 1. CBS and the GR-directed drugs derived from it.

This year, we have been able to determine whether CBS-4 and CBS-5 are selectively cytotoxic to those cells that express GR. Many tumors of the gastrointestinal tract express GR abundantly. This is not the case, however, for cell lines derived from these tumors, because the cells cease to express GR during the first few passages in culture. We have recently obtained a rat pancreatic acinar cell carcinoma line (AR42J) from Robert Jensen (National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD) that expresses GR relatively stably. We have been able to show that gastrin, CBS-4, and CBS-5 bind to the receptor on these cells with about the same avidity as on guinea pig stomach membranes (Figure 2).

Because gastrin binding to GR is known to induce signal transduction, we examined whether the CBS drugs are capable of doing the same thing. AR42J cells were grown in chambered wells and were pretreated with a fluorescent calcium indicator, Fura-2. They were then treated with the CBS drugs or with gastrin at various concentrations, and the calcium influx into the cells was monitored by an Attofluor digitized microscope. Figure 3 shows the effect of CBS-4, which indicates an immediate and reversible increase in intracellular calcium. CBS-5 and gastrin produced similar effects. The pretreatment of the cells with either CBS-4 or CBS-5 blocked the gastrin response,

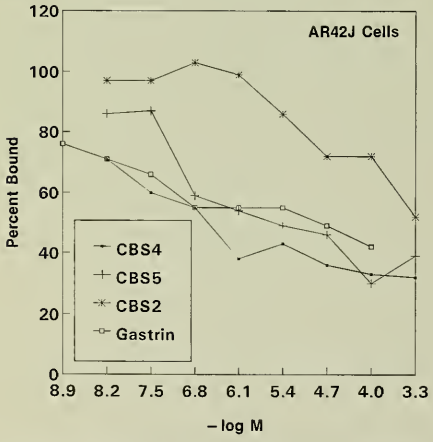


FIGURE 2. GR-binding affinity of CBS compounds and gastrin versus radioiodinated gastrin on AR42J cells.

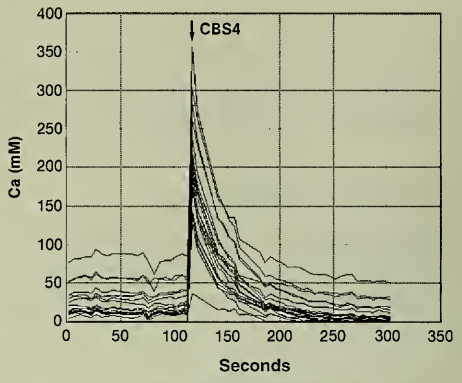


FIGURE 3. Time development of fluorescence from the calcium indicator Fura-2 in AR42J cells treated with the drug CBS-4. The spike indicates the rapid and reversible influx of calcium into the cytoplasm.

suggesting that the effect of the drugs was caused by binding to the same receptor as gastrin. Treatment of the cells with the unconjugated triazene 1-(2-chloroethyl)-3-benzyl-3-(N-methylcarbamoyl)triazene (CBM) caused a large and irreversible calcium accumulation within the cells, suggesting profound toxicity to the cells.

The next question to be answered was whether the CBS drugs, once bound to the receptor, were internalized. We used a rabbit polyclonal antibody to detect CBS-5 and gastrin inside the cells. The cells were incubated with either gastrin or CBS-5 and then fixed with carbodiimide. The cells were then stained with the antigastrin antiserum and detected by fluorescence methods using confocal microscopy. Figure 4 shows that the fluorescent staining produced by gastrin and CBS-5 was very similar, clearly indicating internalization of both the hormone and the drug in the cells, whereas the preimmune serum produced no staining.

We also answered the final key question of selective toxicity. Figure 5 shows the results of a colony survival assay using the AR42J cells. Cells were seeded at low density in 6-well plates. After 24 hours, the drugs at various concentrations were added and the cells were incubated for 24 hours, then the drugs were removed. Gastrin and pentagastrin (the C-terminal pentapeptide) were added every 3 days, and for 2 weeks the cells were allowed to grow into colonies, which were counted after fixing and staining.

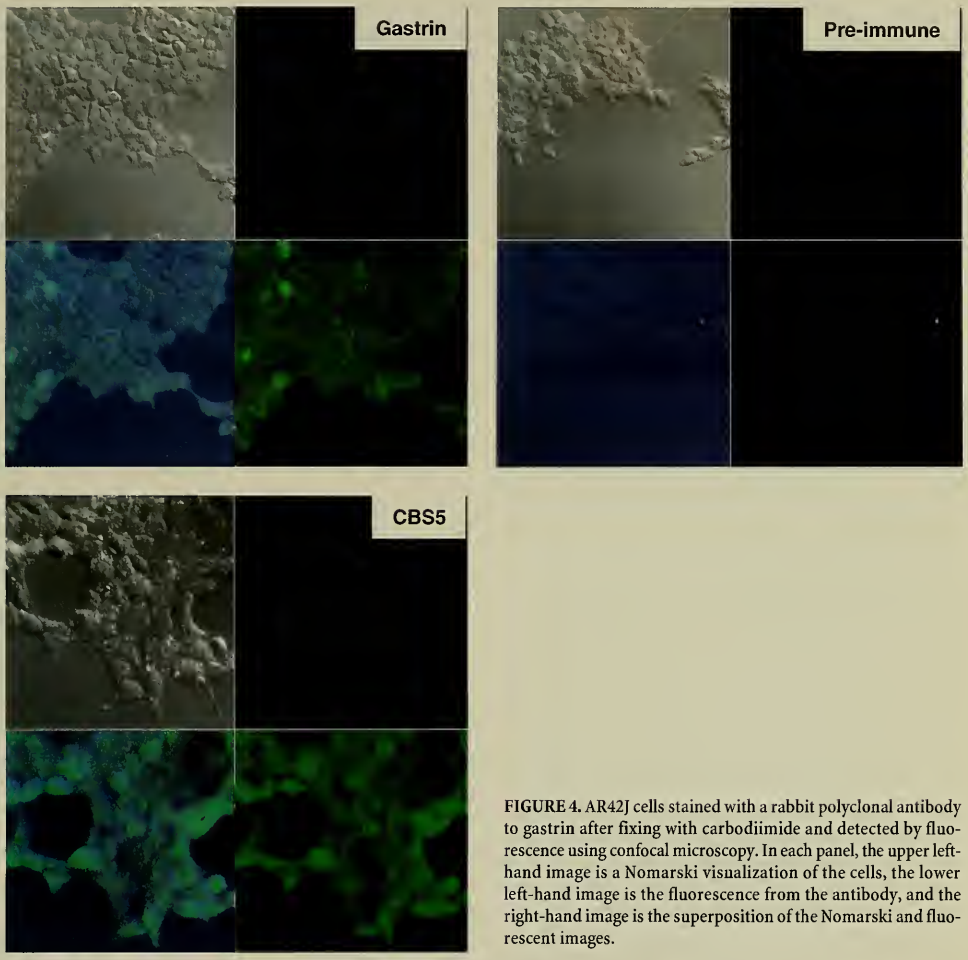


FIGURE 4. AR42J cells stained with a rabbit polyclonal antibody to gastrin after fixing with carbodiimide and detected by fluorescence using confocal microscopy. In each panel, the upper left-hand image is a Nomarski visualization of the cells, the lower left-hand image is the fluorescence from the antibody, and the right-hand image is the superposition of the Nomarski and fluorescent images.

Gastrin and pentagastrin stimulated cell growth by 20% to 50%. Neither CBS-4 nor CBS-2 were toxic to the cells (subsequent data suggested that there were solubility problems with CBS-4). CBS-5, however, was about threefold more toxic than the parent triazene CBM in this assay. This modest gain in toxicity was made more dramatic when the human non-small cell lung adenocarcinoma cell line A549 was used in the assay. A549 cells do not express GR. Treatment with gastrin, CBS-5, and CBM showed that neither gastrin nor CBS-5 had any effect on these cells, while CBM killed them at about the same rate as it did AR42J cells.

These experiments have conclusively shown that the principle of drug targeting via GR is a realistic possibility. We are now in the process of developing other drugs that may show much higher potency yet exhibit good selectivity.

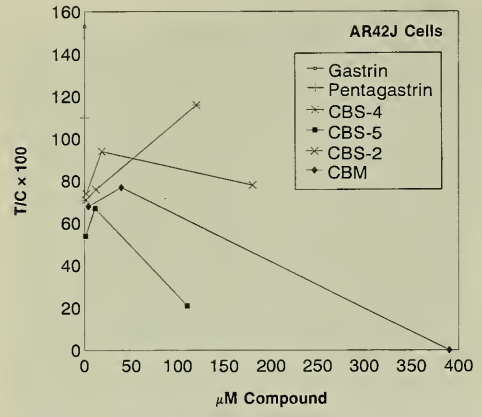


FIGURE 5. Clonogenic assay for cytotoxicity of CBS drugs in AR42J cells

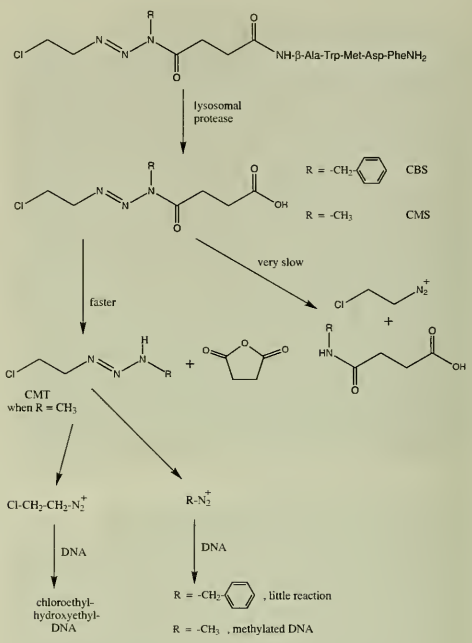


FIGURE 6. Dissociation scheme for CBS-5.

Self-Catalysis in the Decomposition of Acyl Triazenes

The design of drugs such as CBS-5 was predicated on several assumptions that have recently been shown to be correct. It was assumed that CBS-5 would dissociate according to the scheme shown in Figure 6. The proteolytic cleavage of the peptide by cytoplasmic proteases was a reasonable hypothesis, which appeared to be supported by the cytotoxicity data described above. However, the decomposition of the intermediate product CBS to the diazonium ions, which are the putative DNA-damaging agents, was based more on conjecture than on experimental data. Our previous work had shown that the direct dissociation of CBS to the chloroethyldiazonium ion was much too slow to compete with various detoxification mechanisms. However, the synthetic chemists had noted that it was difficult to prevent CBS from decomposing when it was stored as the free acid. Interestingly, CBS was totally inactive as a cytotoxic agent when applied to tumor cells in culture. We assumed that the free acid was not able to penetrate inside the cells because of the polar nature of the carboxyl group. This idea was supported by the finding that the simple methyl ester of CBS was cytotoxic to A549 cells. We assumed that the methyl ester was hydrolyzed by cytoplasmic esterases, and indeed it was found to be an excellent substrate for porcine liver esterase. These experiments showed that, once inside the cell, CBS was cytotoxic; however, we still did not know the mechanism of cytotoxicity.

An indication of the mechanism of activation was provided by the compound CMS, which is identical to CBS with the exception of a methyl group in place of the benzyl group (Figure 6). Our previous work on DNA alkylation by acyltriazenes such as CMC and CMA, under conditions in which the acyl group was removed, showed that these compounds became mainly methylating agents. If the acyl group remained intact, then the DNA was modified by chloroethylation or by hydroxyethylation (which also arises from the chloroethyldiazonium ion). Treatment of calf thymus DNA with CMS methyl ester in buffer in the absence of esterase gave a mixture of methylation and chloroethylation/hydroxyethylation of DNA-guanine. The reaction in buffer is very slow and, hence, DNA methylation observed under these conditions is due to partial hydrolysis of CMS methyl ester to CMS. CMS is mainly a methylating agent

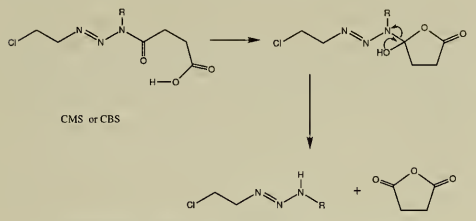


FIGURE 7. Self-catalyzed decomposition of CBS.

under these conditions. In the presence of esterase, however, CMS methyl ester gives no detectable chloroethylation or hydroxyethylation products, consistent with initial hydrolysis to CMS followed by dissociation of CMS to CMT. These data provide conclusive evidence that CMS (and presumably CBS) self-catalyze their decomposition. Molecular modeling suggests that the reaction shown in Figure 7 accounts for the data.

These data support the feasibility of constructing an ester that is a substrate for a tumor-specific protease. Such a protease would release the active metabolite (e.g., CMT) at the proper site of action but would not be able to activate the drug in other tissues. We are actively pursuing this idea.

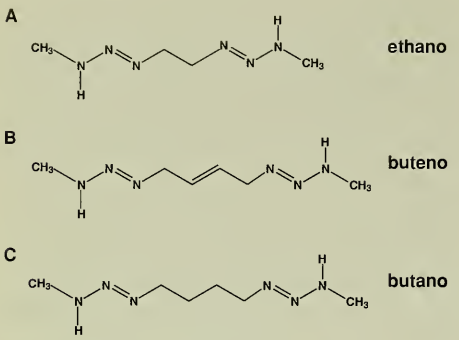
Application of Quantum Mechanical Calculations to Drug Design

Quantum mechanical calculations provide accurate a priori information on the molecular structure and energetics of candidate drugs. The application of these calculations to drug design can be effective in predicting the properties of hypothetical molecules or in explaining the behavior of experimentally known substances.

Quantum mechanical calculations involve the solution, usually at some level of approximation, of the Schrödinger equation ($H\psi = E\psi$). This seemingly simple equation cannot be solved exactly for anything but the simplest atomic and molecular entities. Application of this equation to more complex systems requires the use of various approximation techniques. The most common of the fully theoretical methods involves the so-called Hartree-Fock (HF) approximation, in which the Hamiltonian operator, H, in the above equation is substituted by the approximate Fock operator, F. The solution of the HF equation, which uses the linear combination of atomic orbitals (LCAO) approach to provide information on molecular orbitals, involves the self-consistent field (SCF) method to evaluate the wave function and the energies of a given molecule. Therefore, the calculations start with a predicted structure, whose molecular parameters are then varied using the Berny optimization routine until a structure with a minimum energy is obtained. For large molecules, such as candidate drugs, this approach requires the use of high-speed, large-memory computers. The calculations discussed here were carried out on the CRAY-Y MP-8 supercomputer at the Biomedical Supercomputing Center at the FCRDC.

In practice, these calculations are carried out using various well-established program packages. In these investigations, we used the program GAUSSIAN 92, operating within the CRAY UNICHEM package. The GAUSSIAN program uses the HF approach, in which a wide assortment of basis functions can be chosen. These functions provide an approximate description of the electron configurations used in LCAO. In these calculations, we used the most economical double zeta basis set, 3-21G, to compute the initial structures or reaction paths, which were then repeated using the more advanced basis set, 6-31G*. The final calculations involved the application of the Möller-Plesset (MP) perturbation theory to provide some information on electron correlation effects. We also carried out some of the calculations by using the density-functional method. This relatively new method, in principle, converges to lowest energy structures more rapidly than GAUSSIAN and implicitly includes some contributions of electron correlation. We used the D GAUSS program from CRAY Research for the density-functional calculations.

The great advantage of quantum mechanical calculations is that they can provide a relatively accurate a priori description of the stereo-electronic properties of molecules. Since no experimental information is required, the calculations can be carried out on completely hypothetical substances, thereby providing a reliable prediction of properties. The great disadvantage of quantum mechanical methods is that they are very costly in computer time. The calculations at the HF level increase by approximately N^4, where N is the number of atoms. This increase means that macromolecular calculations by these methods are not practical at present. However, the results on smaller molecules



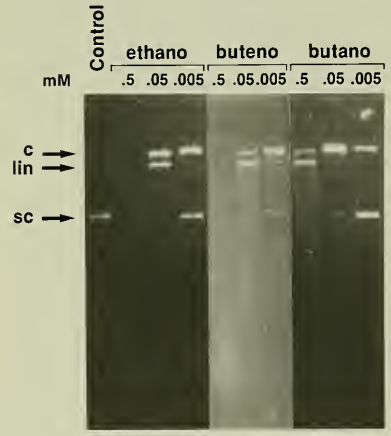


FIGURE 8. Bistriazenes and their reaction with pBR322. Circular (C), linear (Lin), and supercoiled (SC) forms of pBR322 were separated by agarose gel electrophoresis. Untreated pBR322 was included as a control.

can be very useful. Here, we describe two computational projects in which we solved relatively intractable problems.

Figure 8 shows three bistriazene structures. Bistriazenes are a class of compounds that exhibit very powerful cytotoxic properties. One of our studies of bistriazenes involved the measurement of DNA damage, using the supercoiled plasmid pBR322. Figure 8 shows that the ethano- and buteno-bridged bistriazenes A and B are tenfold more reactive toward the plasmid than the butano-bridged bistriazene C. We carried out HF calculations at the 6-31G* basis set level on conformations of the bistriazenes A and C. This basis set level was shown by our previous work to give accurate predictions of structure. In fact, the lowest energy conformation of bistriazene A was almost superimposable on the structure determined experimentally by X-ray crystallography. This structure is shown as A in Figure 9. Note that the structure is an extended helical conformation, which is almost ideally set up for coordination with DNA, as we had indicated by molecular modeling techniques. The most stable conformation of the butano-bridged bistriazene was a folded structure (panel C in Figure 9), which was about 20 kcal/mol more stable than the nearest extended conformation (panel B in Figure 9).

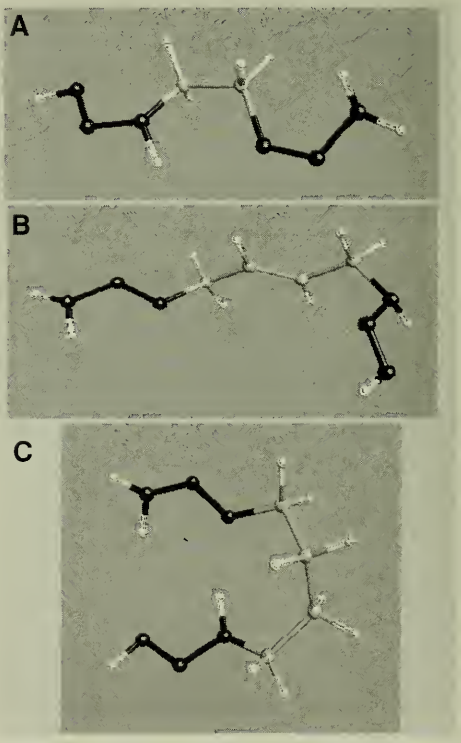


FIGURE 9. Computationally derived conformations of bistriazenes.

This structure would not be expected to fit into the major groove of DNA as well as bistriazene A. The trans-butenobridged bistriazene B is geometrically constrained from folding into the internally hydrogen-bonded structure and, as a result, is at least as active as bistriazene A toward the plasmid.

It has been postulated that one of the intermediates that is formed during the dissociation of chemotherapeutic chloroethylnitrosoureas, such as bis-2-chloroethylnitrosourea (BCNU), is a highly unstable heterocycle 4,5-dihydro-1-oxa-2,3-diazoline (structure 2 in Figure 10) (Cancer Treat Rep 60:651, 1976). This compound would also be expected to form from 2-chloroethyltriazenes, although demonstrating such formation experimentally has proven to be very elusive. We approached this problem computationally by applying quantum mechanical methods to probe the likely reaction pathways for the decomposition of the molecule. These pathways are indicated in Figure 10. We proposed that in the absence of protons, oxadiazoline 2 could dissociate along three possible pathways (A, B, and C), which would proceed through transition states related to products 4, 5, and 6 and on to products 8, 9, and 10.

The highest occupied molecular orbitals (HOMOs) for the starting oxadiazoline and the transition states are shown in Figure 11. We obtained these structures by probing the reaction hypersurface, using the HF/6-31G* level of theory. The ball-and-stick representation of the atoms is accurate with respect to geometry. The "balloons" are regions that define the highest probability of finding the two highest energy electrons for this assembly of atoms. Knowledge of these shapes, especially of the polarities as indicated by the two colors, is very useful in predicting the mechanistic details of the reaction. These calculations were actually carried out at several levels of theory, ranging from the semiempirical PM3 calculations to the highly advanced minimization using the MP perturbation theory at the 6-31G* basis set level (MP2-6-31G*). The calculated energies of activation for each level are shown in Table 1.

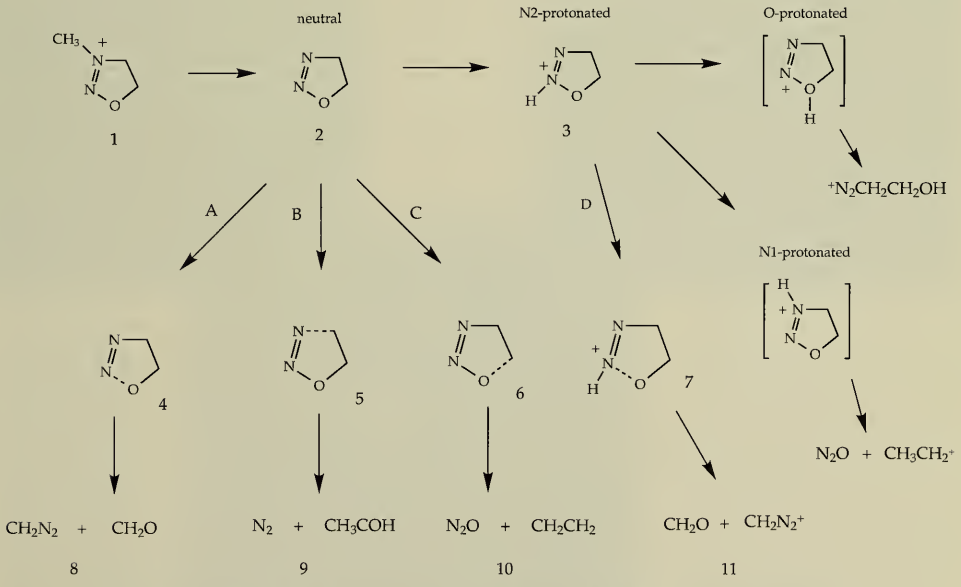


FIGURE 10. Possible reaction pathways for the decomposition of 4,5-dihydro-1-oxa-2,3-oxadiazoline.

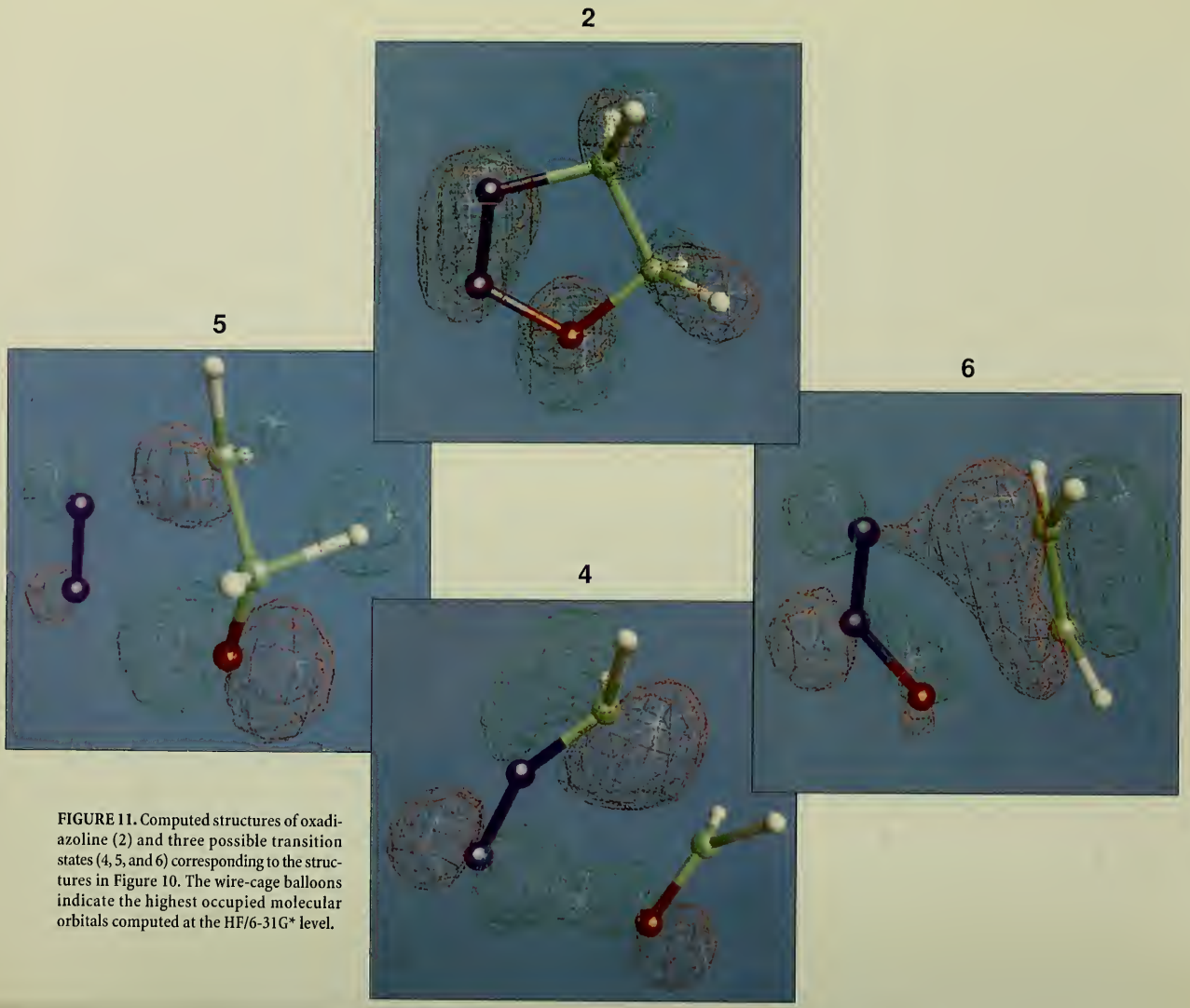


FIGURE 11. Computed structures of oxadiazoline (2) and three possible transition states (4, 5, and 6) corresponding to the structures in Figure 10. The wire-cage balloons indicate the highest occupied molecular orbitals computed at the HF/6-31G* level.

TABLE 1. Energy of activation (kcal/mol) for oxadiazoline bond-breakage at different basis set levels

Bond-breakage pathway	PM3	HF/3-21G	HF/6-31G*	MP2/6-31G*
A	34.5	37.3	27	18.76
B	48.7	45.6	29.2	25.16
C	56	53.9	49.2	28.80
D	NC*	62.8	70.3	40.00

*Not calculated.

At any given level of theory, pathway A was shown to have the lowest activation energy, which strongly suggests that in the absence of protonation, oxadiazoline **2** should decompose to diazomethane. An activation energy of 19 kcal/mol at the highest level of theory suggests that oxadiazoline **2** is not a very stable molecule, as we have confirmed experimentally. These results predict that chloroethylnitrosoureas are methylating agents under some conditions. A recent, still preliminary finding in our experiments indicates that BCNU methylates DNA at pH>8. At pH 7.0, no methylation was evident. We suspect that in neutral pH, there is sufficient H_3O^+ present to O-protonate the oxadiazoline, causing it to decompose via the hydroxyethyldiazonium ion pathway (see Figure 10).

Macromolecular NMR Section

R. Andrew Byrd

Amanda S. Altieri
Donna M. Baldisseri
Paul J. Dowling, Jr.
Kristine E. Miller

The Macromolecular NMR Section was created in 1992 and focuses on the application of modern high-field nuclear magnetic resonance (NMR) techniques and computational methods for the determination of macromolecular structures in solution. The major objective of this section is the elucidation of the three-dimensional (3D) structures of proteins and protein complexes (with drugs or nucleic acids) in solution at atomic resolution as a means of gaining insight into the function of these proteins and the molecular mechanisms of their action. The research areas of interest involve the processes of signal transduction and transcriptional regulation. Such biological events derive from the intermolecular interaction of very specific molecular targets with effector proteins.

Our section has experienced an exciting year with the completion of laboratory space renovations, the delivery and installation of two high-field NMR spectrometers [operating at 500 and 600 MHz; each equipped with four channels and pulsed field gradient (PFG) technology], and the expansion of our staff. During the next contract year, we will be adding two staff members: one NMR scientist and one research technician. The 1992 Annual Report provided a very brief overview of the strengths of the NMR approach to determining macromolecular structures and some of the special requirements, such as isotopic labeling with ^{13}C and ^{15}N. This year, in addition to completing our laboratory setup, we have begun to apply these approaches to a number of protein structure research projects, which will be described below.

There are projects underway to examine the structures of two small proteins without the use of isotopic labeling. These proteins are the neurotoxin anthopleurin B (in collaboration with G. Marius Clore and Angela Gronenborn, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD, and Kenneth M. Blumenthal, University of Cincinnati College of Medicine) and amylin, which has been implicated in Alzheimer's disease (in collaboration with Daruka Mahadevan, National Cancer Institute, Bethesda, MD, and Narmada Thanki, Enzyme Structure Section). Furthermore, we have established collaborations to examine the structure of larger proteins involved in signal transduction and transcriptional regulation. These collaborations include efforts with Monica Tsang (R & D Systems, Minnesota, MN) for the study of the cytokines, such as monocyte chemotactic activating factor (MCAF/MCP-1), with Peter F. Johnson (Eukaryotic Transcriptional Regulation Group, Molecular Mechanisms of Carcinogenesis Laboratory) for the study of the family of transcription-regulating proteins that are related to CCAAT/enhancer-binding protein (C/EBP) and known as C/EBP-related proteins (CRPs), and with George N. Pavlakis and Barbara K. Felber (Human Retrovirus Section and Human Retrovirus Pathogenesis Group, Molecular Mechanisms of Carcinogenesis Laboratory) for the study of the interferon-induced RBP927 protein. Moreover, a collaboration is under discussion with Anna Marie Skalka (Institute for Cancer Research, Fox Chase Cancer Center, Philadelphia, PA) for the study of human immunodeficiency virus (HIV) integrase protein.

In addition, we are involved in the investigation of new methods in both the experimental NMR data acquisition and structure computation phases of the overall structure determination process. The experimental NMR efforts

involve the use of PFG technology to improve the existing protocols for spectral acquisition, assignment, and extraction of structural information. We are using these approaches both for nonlabeled proteins, such as anthopleurin B and amylin, and for larger, isotopically labeled proteins. The advantages of these approaches include increased sensitivity and reduced data acquisition times, both of which have become critical to the study of sensitive or unstable proteins. Our efforts in the computational area involve the integration of new approaches to speed up the initial approximate structure calculation with the complete assignment of all internuclear distance data and high-resolution structure refinement.

Investigation of Cytokine Structure by NMR

Cytokines form a large superfamily of proteins that communicate between different cells of the immune system. The structure-function relationships of these proteins are an intriguing field of study, since members of this family may exhibit a high degree of sequence similarity but have widely varying activities. Using NMR spectroscopy, we are investigating the solution structure of several of these cytokines, including MCAF/MCP-1, in collaboration with the crystallographic efforts of Alexander Wlodawer (Enzyme Structure Section) and the protein expression efforts of Dr. Tsang.

MCAF/MCP-1 is a 76-amino acid polypeptide (8.5 kilodaltons) that exists as a homodimer. MCAF/MCP-1 shares 24% sequence identity with interleukin-8 (IL-8), whose crystal structure was solved by Dr. Wlodawer's section with the assistance of coordinates from the high-resolution NMR structure determined by Drs. Clore and Gronenborn. There exists no crystal or NMR structure for MCAF/MCP-1 at this time; however, a molecular modeling study indicates that its structure should be very similar to that of IL-8. Dr. Wlodawer's section has accumulated diffraction data for MCAF/MCP-1, but attempts to solve the crystal structure using IL-8–derived coordinates have been unsuccessful.

Preliminary results from our section indicate that MCAF/MCP-1 is amenable to study by NMR spectroscopy. The one-dimensional (1D) ^{1}H spectrum (Figure 1) shows reasonable dispersion in the amide proton region as well as the presence of several resolved, upfield-shifted methyl

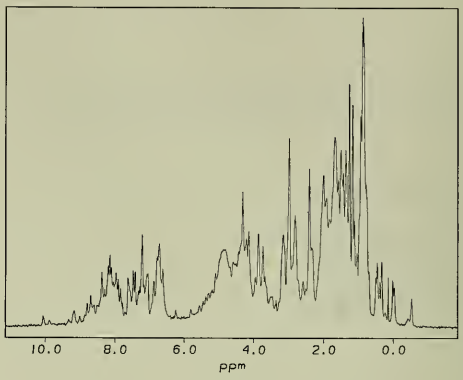


FIGURE 1. ^{1}H spectrum (500 MHz) of MCAF/MCP-1 at 37°C. The protein concentration is 1.4 mM in 90% H_2O/10% D_2O with 10 mM HOAc-d_3 and 0.1 M NaCl at pH 3.0.

proton resonances, which are spectral characteristics that indicate a folded solution structure. Additionally, the two-dimensional (2D) nuclear overhauser effect spectroscopy (NOESY) and total correlation spectroscopy (TOCSY) spectra show many intense correlations that must be measurable for structural analysis. Because of resonance overlap, structure determination of proteins of this size is not tractable by NMR methodologies unless they are labeled with the stable isotopes ^{15}N and ^{13}C. Uniform or selective labeling with one or both of these isotopes allows the application of double- and triple-resonance techniques that greatly simplify the overcrowded spectra obtained from natural-abundance proteins of high molecular weight.

Dr. Tsang's laboratory has provided us with ^{15}N-labeled MCAF/MCP-1, which will facilitate the sequential resonance assignment procedure necessary for subsequent structure determination. This ^{15}N-labeled protein yielded an extremely well-resolved ^{1}H-^{15}N heteronuclear single quantum correlation (HSQC) spectrum (Figure 2). We used new experimental protocols (see below) that enabled the collection of this spectrum in only 30 minutes. These data indicate that the protein is well structured and that the full battery of multidimensional triple-resonance techniques should be readily applicable to this protein.

In addition to MCAF/MCP-1, Dr. Tsang will make several other isotopically labeled cytokines available to us, such

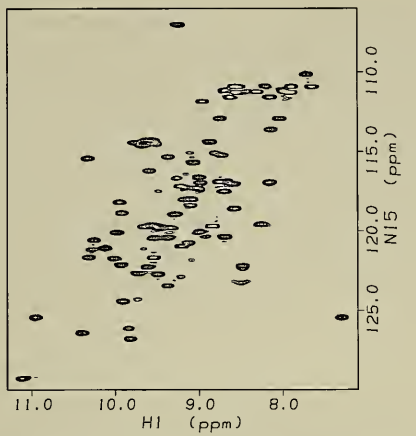


FIGURE 2. ^{1}H-^{15}N 2D HSQC spectrum of 1.5 mM [^{15}N]MCAF at 37°C in 95% H_2O/5% D_2O with 0.1 M NaCl at pH 6.4. The spectrum was acquired using the pulse sequence shown in Figure 5 on a Varian UnityPlus 600 MHz spectrometer in a triple-resonance PFG probe. The spectrum was recorded with only two scans per t_1 increment and no presaturation of the H_2O resonance. Data acquisition was completed in 30 minutes compared to an expected 10 to 12 hours if collected without PFGs and complete phase cycling.

as IL-10, IL-13, and melanoma growth-stimulating activity (MGSA) protein as well as mutants. It is our long-range goal to closely examine the structural relationships between these cytokines. The elucidation of the three-dimensional structures of these proteins should help to clarify the subtleties between structure and function.

Examination of Small Nonlabeled Proteins in H_2O Solutions

Our section is initiating structure determination studies of two small proteins, anthopleurin B (49 amino acids) and amylin (37 amino acids). Both of these proteins have interesting therapeutic potential. Isolated from the sea anemone *Anthopleura xanthogrammica*, anthopleurin B (ApB) is a toxin that binds to fast voltage-dependent sodium channels of nerve and muscle. ApB exhibits a powerful positive inotropic effect, which is correlated to an increase in the amplitude of the action potential and is suppressed by tetrodotoxin. A related protein, anthopleurin A (ApA), has been tested in vivo, and there are indications that it may be more advantageous than glycosides or other inotropic agents in the treatment of heart failure. However, ApB is 7 to 10 times more active than ApA on mammalian cardiac and neuronal channels. There is secondary-structure information for ApA available from NMR. Hence, a complete 3D structural elucidation of ApB compared to ApA could shed light on the mechanistic differences in activity. This type of information is pertinent to the rational design of new therapeutics based on these agents or to utilization of these natural agents themselves.

Amylin makes up the pancreatic islet amyloid associated with noninsulin-dependent diabetes mellitus (NIDDM) in human and rat. This protein is of interest both for its involvement in NIDDM and for the similarities between the islet amyloid and the core particles associated with Alzheimer's disease. The human and rat forms exhibit similar calcitonin-like activity and are potent inhibitors of glycogen synthesis; however, they exhibit differential behavior in amyloid fibril formation. Although a model-building study proposed a structural basis for this behavior, there is no experimentally obtained structure available to evaluate this model. There is also interest in the rat protein for therapeutic action, where the beneficial biological response may be retained without inducement of amyloid formation in humans.

The study of small nonlabeled proteins, such as ApB and amylin, can be accomplished via the traditional ^{1}H homonuclear methods employing 2D correlation spectroscopy (COSY), TOCSY, and NOESY NMR spectroscopic methods. However, because these experiments rely on correlation of labile NH protons with CαH protons and other side-chain protons, all of these experimental protocols must be performed in 90% H_2O solution. This situation creates a relative dynamic range problem of approximately $1:1 \times 10^5$, when one considers the intensity of a typical 1 mM protein signal in the presence of 90% H_2O (~100 M in ^{1}H). The experimental protocol requires efficient suppression of the very intense H_2O resonance in the ^{1}H spectra. In order to observe the protein ^{1}H spectrum and perform the necessary 2D correlation experiments, it is necessary to suppress the intense H_2O resonance by roughly 10^4-fold. Although the methods for achieving water suppression have been an area of active research for the past

15 years, the current approaches still have several problems. When the protein is not isotopically labeled with ^{15}N or ^{13}C, the suppression does not benefit from any heteronuclear filtering that may be applied using a combination of ^{1}H and ^{13}C or ^{15}N nuclei and thus differentiating against H_2O. We have been examining the use of PFGs in combination with new shaped radio-frequency (RF) pulse methods as well as traditional hard-pulse water suppression techniques to achieve improved performance. The shaped pulse methods include both selective excitation of H_2O followed by a crusher PFG and nonexcitation of H_2O in combination with PFGs. The result of this approach both provides better suppression of water and removes differential perturbation of protein resonances due to saturation transfer effects.

The most common traditional approach to water suppression involves selective irradiation of the H_2O resonance, referred to as presaturation. This irradiation has two well-known detrimental side effects: those protons that exchange with the H_2O resonance can be saturated or suppressed, thereby exhibiting weak intensity, or can be lost altogether; and the ability to observe specific interactions with bound H_2O molecules within the protein structure is virtually eliminated. However, there is a third, less-recognized effect that is also very serious. The transfer of saturation from the H_2O resonance primarily affects the labile backbone NH protons; however, this transfer can also affect all of the resonances in the protein, via spin diffusion, thus reducing the effective signal strength for all protons in all experiments. The consequence is an effective reduction in overall sensitivity. These effects exhibit a pH dependence that follows the pKa of the amide protons. The exchange is slower at acidic pH; hence, most protein NMR studies were previously conducted below pH 6.0. The loss of signal intensity with presaturation becomes quite dramatic at neutral pH.

An example of these effects is seen in Figure 3, which shows a comparison of 1D ^{1}H spectra of the protein amylin in 90% H_2O. The lower spectrum was obtained with the traditional presaturation technique and the upper spectrum was obtained using a shaped pulse technique that avoids excitation of the H_2O resonance. The aromatic protons of Phe15 and Tyr37 are indicated. These protons are unaffected to first order by presaturation. Using these protons as an intensity standard, we observed that the

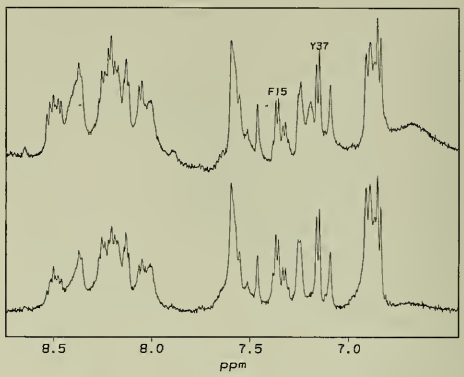


FIGURE 3. Effect of H_2O suppression techniques on the amide region of the 37-amino acid peptide amylin. The bottom spectrum was acquired using presaturation of the H_2O resonance, which reduces the intensity of the exchangeable amide NH protons compared to the ring protons of F15 and Y37. The top spectrum was acquired using a shaped pulse that provides zero excitation at the H_2O frequency and shows an increase in intensity of the exchangeable NH protons. The spectra were acquired on a Varian UnityPlus 500 MHz spectrometer.

NH region of the spectrum (7.8 to 8.8 ppm) clearly exhibits 40% to 70% increased signal intensity in the shaped pulse spectrum.

By taking full advantage of the latest instrumental technology, we have combined the shaped pulse methodology with PFG methodology to develop variations of the traditional 2D ^{1}H experiments that can be implemented without the need for any irradiation of the H_2O resonance and without any undue complications or variation in data processing or analysis. Our initial results indicate an increase in intensity of protein amide resonances by up to 70% compared to conventional presaturation, especially at neutral pH. In addition, the use of PFG-assisted techniques leads to better spectral fidelity and minimizes postacquisition processing of the data. This approach will provide more precise interproton distance measurements and should lead to more accurate 3D structures. An example of this approach is seen in Figure 4, which is a 2D NOESY spectrum of the protein ubiquitin obtained without presaturation of the H_2O resonance. An optimized shaped RF pulse is used in the final detection segment of the experiment

and PFGs are used in the NOE mixing time. The dashed line in the contour plot indicates the frequency of the H_2O resonance. A trace through the 2D spectrum at this H_2O frequency is seen at the top. The peaks along this trace indicate the protein protons that show interactions, either via chemical exchange or NOE, with H_2O. Analysis of these interactions can identify tightly bound H_2O molecules within the protein structure. Previous methods for observation of these interactions have been considerably more arduous and often required isotopic labeling. Our approach is straightforward and may be applied to small proteins without the need for isotopic labeling. We anticipate that we will have appropriate variants of the standard 2D correlation experiments to enable a complete study of ApB and amylin without presaturation effects.

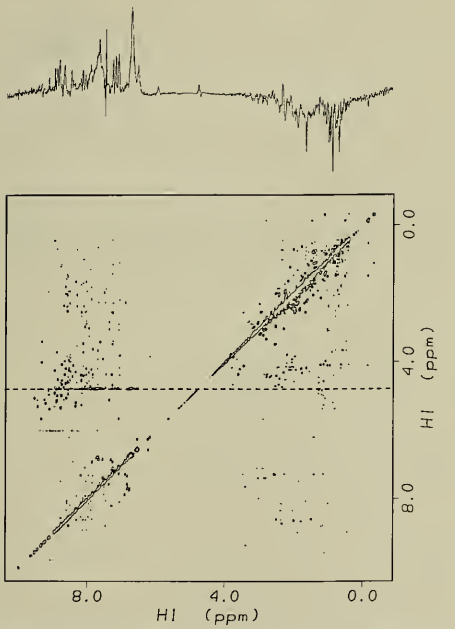


FIGURE 4. 2D NOESY spectrum (500 MHz) at 30°C of ubiquitin (76 amino acids) at a concentration of 2 mM in 90% H_2O/10% D_2O with 10 mM HOAc-d_3 at pH 5.5. No presaturation of the H_2O resonance was used, and the pulse sequence incorporated a 3 msec 20 G/cm PFG in the NOE mixing period and an optimized shaped RF pulse as a read pulse. This read pulse resulted in an inversion of sign of intensity for resonances upfield of H_2O as compared to downfield of H_2O. The dotted line is shown at the ^1H frequency of the H_2O resonance. The trace through the 2D spectrum at this point is shown at the top of the figure.

Combined Use of PFGs and Isotopic Labeling in Resonance Assignment by Multidimensional Double- and Triple-Resonance NMR Spectroscopy

The combination of isotopic labeling with ^{13}C and ^{15}N and the use of multidimensional double- and triple-resonance NMR experiments has revolutionized the process of spectral assignment and structure determination of medium-sized proteins via NMR spectroscopy. However, there are still significant problems in terms of the quantities of material required (ca. 1 to 2 µmol) and the stability required (thermally stable at 37 °C for several periods of 3 to 10 days each) to complete the data collection phase. One of the causes of the stability requirement is the extensive cycling of RF pulse phase within a given RF pulse sequence in order to select the particular spectral component desired. This phase cycling often leads to minimal data acquisition times for the basic experiment that are in excess of the required signal averaging for sensitivity. For example, 3D triple-resonance experiments normally require 2 to 3 days of acquisition to accomplish the appropriate spectral selection via phase cycling and to provide minimal digital resolution in all three dimensions. The problem is exacerbated when the experiment is extended to four dimensions. There has been a considerable effort in recent years to address this problem via the introduction of PFGs to select specific spectral components, e.g., coherence selection. However, not unexpectedly, this coherence selection leads to problems of its own, namely either the loss of pure absorption mode signals or more complex acquisition and processing procedures to maintain sensitivity. It has recently been demonstrated that PFGs may be used to radically reduce or eliminate conventional phase cycling (Bax and Pochapsky, J Magn Reson 99:638, 1992) without increasing the complexity of acquisition or processing. This use of PFGs can result in greatly reduced minimum acquisition times.

We have been incorporating these concepts into numerous double- and triple-resonance experiments. A comparison between the non-PFG and PFG-based experiment

for ^{1}H-^{15}N correlation is shown in Figure 5. The minimum phase cycling is reduced from 16 or 32 steps to 2 steps. This reduction in phase cycling correlates directly to a reduction in experiment time, provided that sufficient H_2O suppression is also maintained. In fact, this sequence provides excellent performance for both aspects and permits rapid recording of protein ^{1}H-^{15}N correlation spectra, as shown in Figure 2. This spectrum was recorded in only 30 minutes, and it is possible to reduce this time even further by making more extensive use of advanced data processing procedures, such as linear prediction. Clearly, by incorporating this approach into 3D experiments, it is anticipated that the minimum acquisition times for some 3D correlation experiments will be reduced to approximately 8 hours or less. An example of the complexity of such sequences is given in Figure 6. These efforts follow our similar developments mentioned above and will permit our section to deal with less well behaved proteins and to maintain optimal performance.

The time comparisons mentioned above have assumed the preferred operating conditions for NMR experiments of approximately 1.0 mM protein concentration in a volume of 0.5 ml in a 5-mm diameter NMR tube.

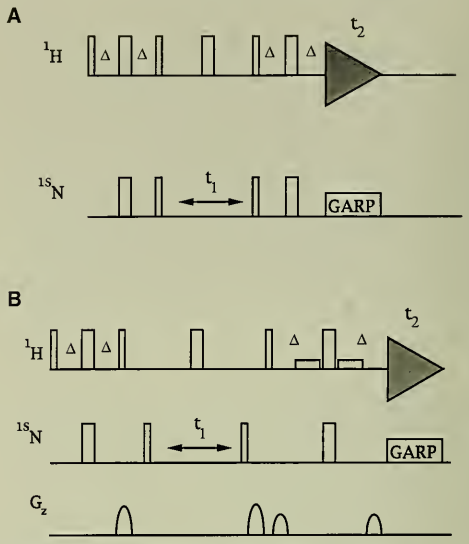


FIGURE 5. RF pulse sequences used for a conventional HSQC experiment (A) and a modified HSQC experiment incorporating PFGs to minimize phase cycling and suppress H_2O (B).

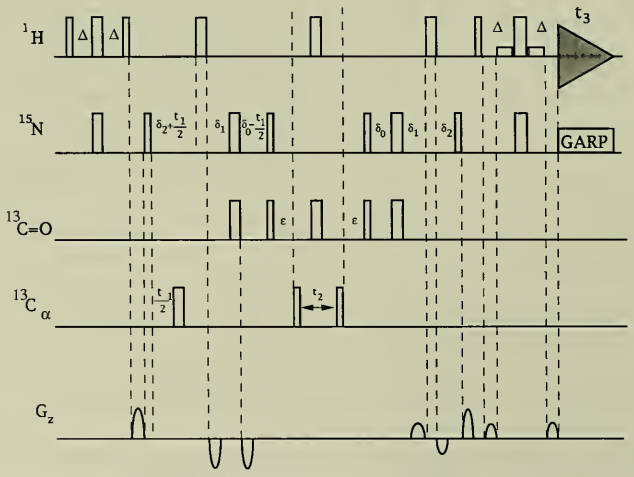


FIGURE 6. RF pulse sequence for a triple-resonance, constant-time HN(CO)CA experiment. PFGs were used to minimize phase cycling and suppress H_2O.

Often it is very difficult to achieve these conditions, just as it is often difficult to determine the conditions under which a protein will crystallize. Consequently, we have initiated a collaboration with Toby Zens (Nalorac Cryogenics Corporation, Martinez, CA) to participate in the development of NMR probes capable of using larger sample volumes and incorporating the requisite triple resonance and PFG features. This equipment will effectively cut in half the required concentrations for NMR studies, but it will not reduce the total amount of protein required. Nevertheless, the reduced concentration may prove to be essential to the investigation of numerous important proteins.

Increased Efficiency in Structure Calculations and Integration with Resonance Assignment

The main target of the NMR technology referred to above is the elucidation of structural information for the subject protein or protein-nucleic acid complex, which is most commonly thought of in terms of a high-resolution solution structure similar to an X-ray structure. The full range of NMR technology is required, since the NMR method utilizes a large number of short-range distance measurements to subsequently calculate a 3D structure that is consistent with the observed measurements. In order to properly measure this large number of distances, it is essential to have correct resonance assignments and unambiguous identification of all NOE interactions. The former condition is much more easily satisfied than the latter. The general approach is outlined in Figure 7. The quality of the final structure is directly proportional to the number of distances that can be detected and unambiguously assigned. The most important distances are those between amino acid residues that are separated in the linear sequence by more than three residues, since these distances report on the spatial arrangement of the backbone fold. A key element is the central iterative cycle of the procedure in Figure 7, which utilizes the evolving 3D fold of the protein to confirm and increase the assignments of long-range NOE interactions. The calculation of a 3D structure via distance geometry or the hybrid approach, as part of this loop, is dramatically impacted by the number of atoms in the molecule. Therefore, as the system under study becomes larger, in the range of 120 to 200 amino acids, the

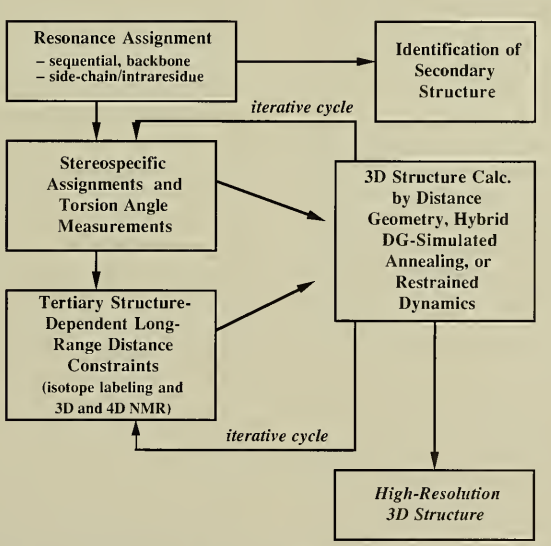


FIGURE 7. General outline of the stepwise procedure for structure determination by NMR and computational methods.

process to calculate a new 3D structure for confirming and expanding the NOE assignments becomes slower. Hence, there is a need for an interactive approach that provides rapid determination of the approximate protein fold to be used to verify assignment of experimental NOEs and establishment of new distance constraints to more accurately define the structure.

We have initiated a procedure within our existing distance geometry software, Dspace (Biosym Technologies, Inc.), that reduces the effective number of atoms to two per amino acid residue. The experimental constraints are adapted to this two-particle representation and a normal refinement protocol is followed. The subsequent structure can be transformed to the all-atom version using commercial modeling software (Insight II, Biosym Technologies, Inc.) and then utilized in the interactive analysis of NOE data. This procedure follows the work of Hoch and Stern (J Biomol NMR 2:535, 1992). Dr. Altieri has confirmed that approximate folds of the protein backbone may be obtained via this approach with as much as a tenfold reduction in time. This approach should accelerate the extraction of all NOE data from the experimental spectra and provide for complete data sets to support the final structural refinement using simulated annealing and restrained molecular dynamics.

C/EBP-Related Transcription Factors

As mentioned previously, we have initiated a collaboration with Dr. Johnson to determine the structure of members of the class of DNA-binding eukaryotic transcription factors that are related to C/EBP. Dr. Johnson's group has been studying these proteins extensively and has had an ongoing collaboration with the Macromolecular Structure Laboratory to determine the structures by crystallographic methods. As a result of this previous work, the protein CRP2, and smaller constructs, have been cloned and expressed at high levels in *Escherichia coli*. Typical yields are 10 to 15 mg/L of culture. This expression level is sufficient to permit efficient isotopic labeling with both ^{13}C and ^{15}N. The protein is known to dimerize and to bind to duplex oligodeoxyribonucleotides (oligoDNA) that are 16 to 20 base pairs in length. We are currently working to determine solution conditions appropriate for NMR study of the protein free of oligoDNA, which is assumed to be a dimer. This determination entails dealing with solubility problems and attempting to minimize salt concentrations that degrade the sensitivity of the NMR instrumentation. In the event that the uncomplexed protein cannot be stabilized at sufficient concentrations for NMR studies, we will proceed with the protein-oligoDNA complex. The total size of this complex (38 to 40 kilodaltons) is approaching the limits of feasibility for NMR technology; however, we are working with Dr. Johnson to examine smaller constructs.

Interferon-Induced Protein RBP927

Drs. Pavlakis and Felber have identified, cloned, and expressed in *E. coli* the protein RBP927, which is induced in response to interferon treatment and has been shown to inhibit HIV type 1 (HIV-1) expression. We have initiated a collaboration to investigate the solution structure of this protein in order to gain some insight into its mechanism of action. RBP927 is known to bind tightly to the Rev-responsive element (RRE); however, the 300-base RRE ribonucleic acid bound to RBP927 represents a complex much larger than NMR technology can deal with. The protein itself is medium sized (126 amino acids, 13.9 kilodaltons) and should be amenable to NMR study. We are investigating the proper solution conditions and hope to proceed promptly to isotopic labeling. We anticipate that some further work may be required on the expression system to enable dual ^{13}C and ^{15}N labeling of this protein.

Molecular Pharmacology Group

Ronald Rubin

Avis Danishefsky
Maria Miller
Julie A. Pavlovsky
April Rollins
Cleopas Samudzi
Bangalore K. Sathyanarayana
Jian-Cheng Xuan

The Molecular Pharmacology Group is investigating the structure and function of proteins that are involved in intercellular communication and signal transduction. In the area of cellular communication, and in particular the immune response, we are determining the structures of γ-interferons (IFN-γ) from several different species, including the human, rabbit, cow, and mouse isoforms of IFN-γ. In the past year, we have determined the crystal structure of recombinant bovine IFN-γ as well as a new crystal form of rabbit IFN-γ. We have refined our structure of rabbit IFN-γ at high resolution and will compare this structure with the structures of IFN-γ from other species. We have also initiated crystallization studies of a complex of murine IFN-γ with the extracellular ligand-binding domain of the murine IFN-γ receptor. In addition, we have begun crystallization trials on human IFN-αA as well as canine and murine IFN-γ.

We are continuing our structural studies on bacterial enterotoxin "superantigens," such as enterotoxin A from *Staphylococcus aureus*. This molecule is able to subvert the immune system and induce a catastrophic immune response, such as that observed in toxic shock syndrome.

A new area of interest for our group is in the related field of autoimmune diseases such as multiple sclerosis (MS). A primary epitope responsible for the autoimmune response observed in MS is myelin basic protein (MBP), which is a major component of the myelin sheath that insulates the neural axons in the central nervous system. We have initiated crystallization and model-building studies of MBP to better understand its role in the induction and progression of MS and to find methods to block the T-cell response to MBP.

In the area of intracellular signal transduction, we are investigating the structure and regulation of GTP-binding proteins related to the *ras* oncogene product $p21^{ras}$. In addition, we are continuing our research on the structure and function of oncogene protein kinase domains such as those from the *mos*, *raf*, and *trk* proto-oncogene products.

We are also continuing our crystallographic studies of transcription factors related to the CCAAT/enhancer-binding protein (C/EBP) to determine how these molecules are regulated and how they recognize specific DNA sequences.

In a joint project with Alexander Wlodawer and the Enzyme Structure Section, we are examining the crystal structure of a synthetic unnatural D-enantiomer of the human immunodeficiency virus type 1 protease (HIV-1 PR) complexed to a D-hexapeptide inhibitor molecule. This study should contribute to our understanding of the role of chemical chirality in biological systems.

The Structure and Function of Interferons

Using the multiple isomorphous replacement method in conjunction with solvent flattening, we have recently determined the crystal structure of recombinant bovine IFN-γ. The crystals used in this structure determination are orthorhombic, space group $P2_12_12_1$, with unit cell dimensions a = 42.8 Å, b = 79.9 Å, c = 85.4 Å. The structure contains an IFN-γ dimer in the asymmetric unit of structure.

Heavy-atom derivatives were obtained by soaking the crystals in solutions of K_2HgI_4 and cis-$Pt(NH_3)_2Cl_2$. By interpretation of 3.0 Å resolution electron density maps, we determined the bovine IFN-γ structure and refined it to a crystallographic R-factor of 19.2%.

The independent structure solution of the bovine IFN-γ allows us to compare its structure with that of the homologous rabbit IFN-γ structure (Figure 1). Although there is only 65% identity between the sequences of the two proteins, the structures of bovine and rabbit IFN-γ are remarkably similar. Like the rabbit IFN-γ structure, bovine IFN-γ is comprised of six α-helical segments as shown in Figure 2. Each of the structural domains of the IFN-γ dimer contains a four-α-helical core structure. A similar four-α-helical core has been observed in other cytokines: for example, β-interferon (IFN-β), human growth hormone (HGH), interleukin-4 (IL-4), and interleukin-2 (IL-2). In addition, all of these molecules share a common functional role in binding to a cell-surface receptor. By superimposing common structural elements of IFN-γ with HGH, IL-4, and IFN-β (Figure 3), we were able to identify common functional elements such as potential receptor-binding surfaces. For example, we were able to superimpose the IFN-γ structure onto the HGH portion of the known structure of the HGH-receptor complex. Superposition allowed us to propose a model for the structure of the IFN-γ–receptor complex (Figure 4). To verify this hypothetical structure, we are attempting to cocrystallize murine IFN-γ with its extracellular receptor domain in collaboration with Sidney Pestka (Robert Wood Johnson School of Medicine, Piscataway, NJ).

FIGURE 1. A comparison of the α-carbon backbones of the bovine (solid lines) and rabbit (dashed lines) IFN-γ structures.

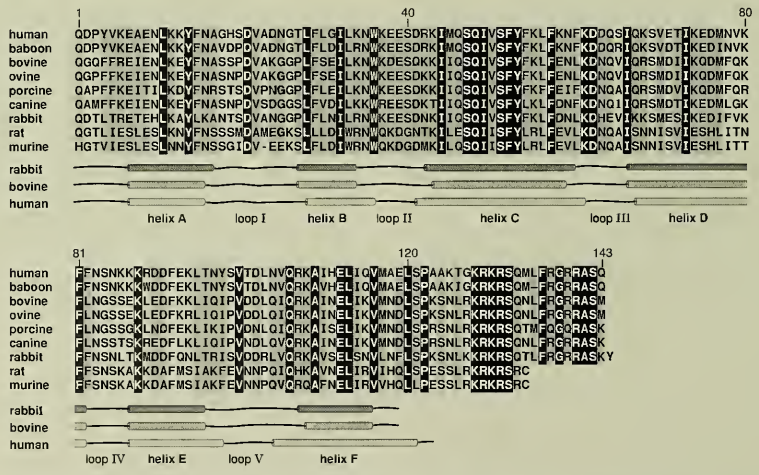


FIGURE 2. An alignment and comparison of the amino acid sequences of IFN-γ from several different species. The tubes below the sequences represent the positions of α-helices in the crystal structures of rabbit, bovine, and human IFN-γ.

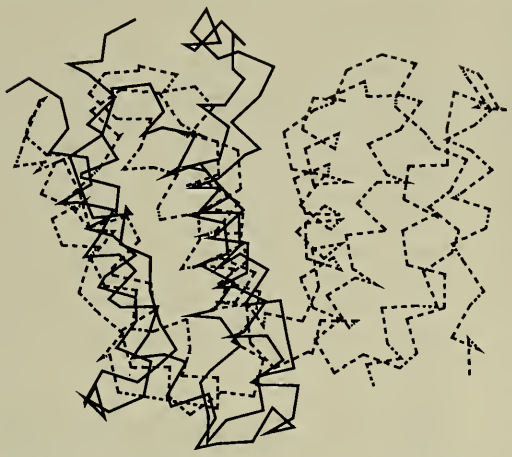


FIGURE 3. A superposition of the Cα structure of the IFN-β monomer (solid lines) onto the rabbit IFN-γ dimer.

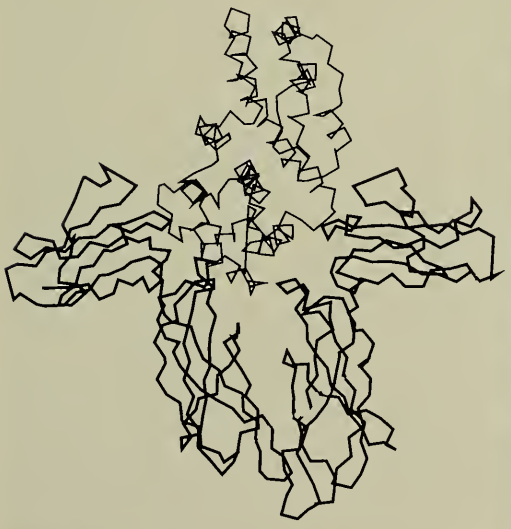


FIGURE 4. The Cα trace of a model for the structure of the IFN-γ–IFN-γ receptor complex. Two receptor domains (bold lines) are shown complexed to a single IFN-γ dimer.

The Structure and Function of Superantigens

In collaboration with Robert Ulrich (U.S. Army Medical Research Institute for Infectious Diseases, Frederick, MD), we are continuing our crystallographic studies of enterotoxin A from *S. aureus* (SEA) as well as site-directed mutants of SEA with altered biological function. We have succeeded in producing small crystals of SEA and two of the SEA mutants (Figure 5), which may be suitable for structure determination by X-ray diffraction methods. The best crystals were obtained by the vapor-diffusion "hanging drop" method at pH 8.5 in 0.1 M Tris buffer and 1.0 M ammonium sulfate. We are currently characterizing this crystal form for further analysis by X-ray diffraction methods.

The Structure and Function of G-Proteins

In collaboration with Donald Court (Molecular Control and Genetics Section, Laboratory of Chromosome Biology), we are studying the structure and function of a class of prokaryotic GTP-binding proteins

FIGURE 5. Crystals of a mutant form of SEA. The crystals are small polyhedrons with maximum dimensions of 0.1 mm.

that are closely related to the p21^{ras} oncoprotein found in mammalian cells. The *Escherichia coli* homolog is known as *E. coli ras* equivalent (ERA) and is required for cell growth in *E. coli*. The sequences of ERA proteins from *Bacillus subtilus* and coccidia have also been determined recently. All three sequences are homologous to other GTP-binding proteins, in particular the GTP-binding domain of p21^{ras}. We have used homology modeling to predict the three-dimensional structures of the GTP-binding domains of ERA proteins from *E. coli*, coccidia, and *B. subtilus* (Figure 6). Dr. Court has isolated a series of mutant ERA proteins with altered activity, and we have located the positions of these mutations on our model. As expected, most of the mutations are in positions close to the GTP-binding site.

To verify the accuracy of our modeled ERA structure, we are purifying large quantities of recombinant ERA that has been overexpressed in *E. coli*. To express ERA, the cells were grown at 32°C and induced at 39°C. The cells were then harvested and lysed by sonication. Soluble ERA was precipitated by addition of ammonium sulfate to the crude cell supernatant to a concentration of 20% saturation. The crude ERA protein was then chromatographed by fast protein liquid chromatography (FPLC) on a mono-Q column to produce ERA with greater than 95% purity. We are currently using this material in cocrystallization trials with GDP and GTP.

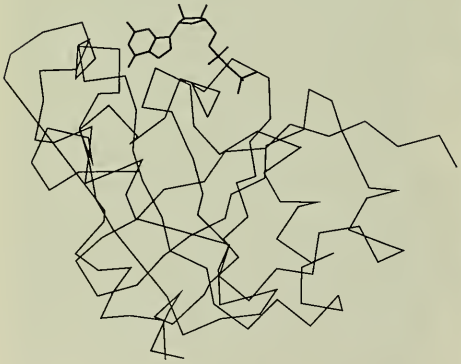


FIGURE 6. A Cα trace of the modeled structure of the GDP-binding domain of ERA, showing the position of bound GDP.

MBP Studies

We have recently initiated crystallographic studies of MBP from bovine brain. The bovine homolog of MBP is a highly basic protein consisting of 170 amino acid residues. We obtained crude MBP from Sigma and we further purified it by FPLC on a mono-S column. We are using this purified MBP in crystallization trials. Small crystals of the protein were obtained in "hanging drops" under a variety of conditions (Figure 7). The best crystals were obtained in 0.1 M ammonium sulfate and 15% (v/v) polyethylene glycol 8000. We are continuing these studies in order to obtain crystals that are suitable for X-ray structure determination.

FIGURE 7. Crystals of MBP obtained from ammonium sulfate–PEG-8000 solutions at 6°C.

Structural Studies of Transcription Factors

We are continuing collaboration with Peter Johnson (Eukaryotic Transcriptional Regulation Group, Molecular Mechanisms of Carcinogenesis Laboratory) on the structure and function of DNA-binding eukaryotic transcription factors. We have purified and characterized large quantities of a deletion mutant of a murine transcription factor (CRP2) that is closely related to C/EBP. We are currently concentrating on a deletion mutant of CRP2, designed by Dr. Johnson, that contains the regulatory and DNA-binding domains with a deletion of about 50 amino acids from the spacer region linking these two domains. When overexpressed in *E. coli*, this mutant protein is produced in temperature-induced cells as insoluble inclusion bodies. The inclusion bodies are harvested from lysed cells by centrifugation, washed, and solubilized in 8 M urea. The recombinant protein contains a $(histidine)_6$ tail at the N terminus and is readily purified by adsorption on a Ni^{3+} chelate column. The column is washed extensively with 8 M urea buffer at pH 8.0 to remove contaminating proteins. CRP2 protein at >95% purity is eluted in 8 M urea at pH 4.5. CRP2 is then refolded by dialysis into refolding buffer containing 0.1 M Tris (pH 9.0), 1.0 M NaCl, 0.1% mercaptoethanol (BME), and 10% glycerol followed by dialysis into 0.1 M Tris (pH 7.5), 1.0 M NaCl, 0.1% BME. This material was assayed by Dr. Johnson's group and found to bind DNA oligonucleotides tightly in a sequence-specific manner. We are setting up crystallization trials of this CRP2 protein, both as the apoprotein and as the complex with several self-complementary oligonucleotides including the 19-mer TATATATTGCGCAATATAT.

The Structure and Function of Oncogene Kinases

On the basis of the three-dimensional structure of the cyclic AMP-dependent protein kinase (cAPK), we have modeled the structures of the kinase domains from several oncogene products, including those from the *raf*, *mos*, and *trk* proto-oncogenes. The accuracy of the modeled kinase structures was tested using the 3D PROFILE program of David Eisenberg (University of California at Los Angeles). The modeled structures were used to make various predictions for each of the kinases, including the possible amino acid recognition sequences surrounding the sites of phosphorylation of protein substrates.

The modeled structure of the kinase domain of the Mos oncoprotein (Figure 8) consists of two domains. The N-terminal ATP-binding domain consists of residues 33 to 143 and contains a five-strand β-sheet structure. The C-terminal peptide-binding domain consists of residues 144 to 346 and is primarily α-helical. The active site of the enzyme occupies a deep cleft between the domains.

George Vande Woude and his co-workers (Molecular Oncology Section, Molecular Mechanisms of Carcinogenesis Laboratory) have shown that Mos protein binds to DNA in the presence of Mg^{2+}-ATP. On one surface of the ATP-binding domain of the modeled Mos kinase structure, there is a cluster of basic side chains including residues

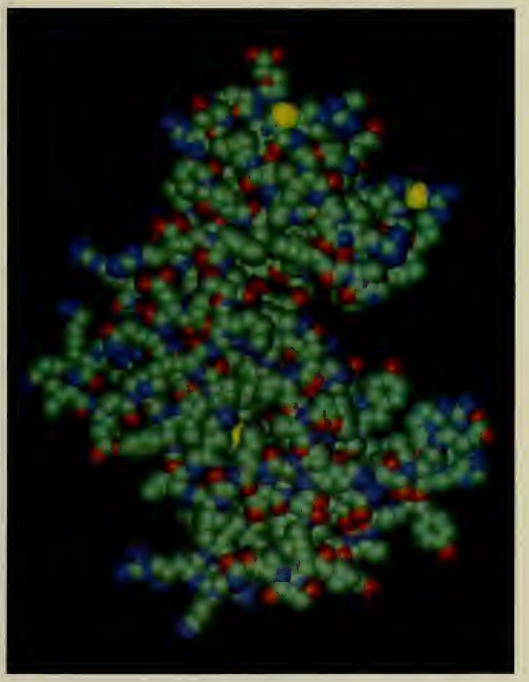


FIGURE 8. A space-filling model of the structure of the kinase domain of the Mos protein.

R43, R46, R49, R50, R81, R112, R114, R120, K128, K181, and K217. These side chains are positioned in such a way that they can interact with adjacent phosphodiester groups along the backbone of double-helical DNA (Figure 9). This configuration of DNA-binding residues appears to be stabilized by the binding of Mg^{+2}-ATP to the opposite side of the ATP-binding domain.

In collaboration with Dr. Vande Woude's section, we are purifying large quantities of recombinant Mos protein for crystallization studies. Mos was expressed as a fusion protein together with maltose-binding protein and purified by affinity chromatography on amylose and Ni^{3+} chelate columns. Attempts to crystallize the purified Mos protein are currently underway.

The Trk receptor tyrosine kinase was recently identified by Luis Parada (Molecular Embryology Section, Mammalian Genetics Laboratory) and David Kaplan (Eukaryotic Signal Transduction Group, Molecular Mechanisms of Carcinogenesis Laboratory) as the major receptor for nerve growth factor (NGF). Dr. Kaplan has identified a number of signal transduction proteins that are substrates for the Trk protein kinase, including the Trk kinase itself. Based on the cAPK structure, we have modeled the Trk kinase domain from residues 531 to 811. Our analysis of the substrate-recognition cleft in Trk kinase suggests that the enzyme should recognize the tetrapeptide sequence RDxY. The Trk kinase contains this motif in the sequence 691 [DFGMRDIY] 700. The phosphate acceptor tyrosine at position 699 is located on a loop on the surface of the Trk kinase domain (Figure 10A). We predict that this tyrosine should be a site for Trk autophosphorylation and have proposed a model for Trk kinase self-association and autophosphorylation (Figure 10B).

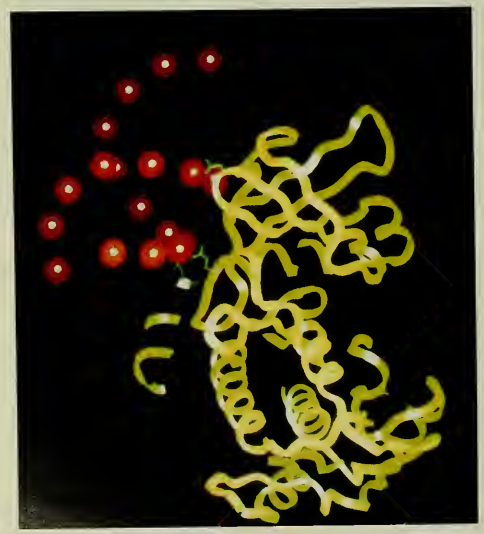


FIGURE 9. A model for the interaction of double-helical DNA with the kinase domain of Mos. The red spheres represent DNA phosphodiester groups.

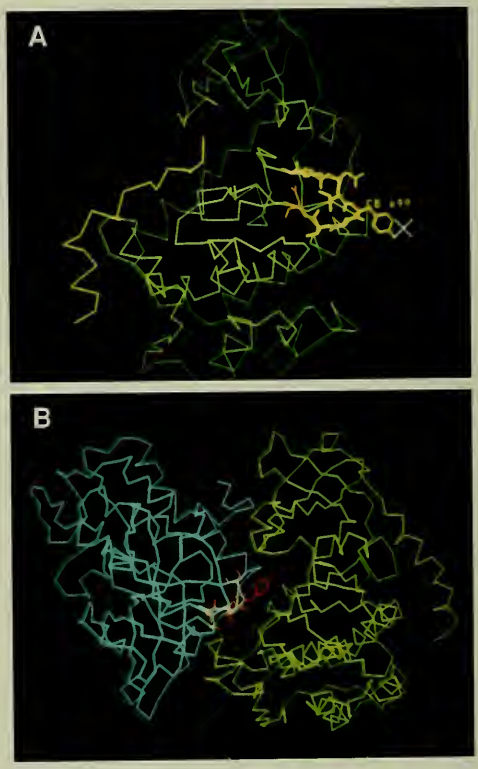


FIGURE 10. (A) A model for the kinase domain of the Trk oncoprotein showing the position of the Tyr699 phosphorylation site. (B) A model for the association of two Trk kinase domains during kinase autophosphorylation.

Crystal Structure of a Synthetic D-Enzyme Prepared by Chemical Ligation

It has been well established that biological macromolecules consist of monomers of uniform chirality and that the recognition and interactions of biological macromolecules require stereochemical specificity. Spontaneous racemization due to aging occurs in tissues such as lens and tooth enamel with detrimental biological consequences. Although D-amino acids were found in antibiotic peptides, these peptides are synthesized in microorganisms via a nonribosomal pathway, catalyzed by multienzymatic complexes.

The recent discovery of the presence of D-amino acids in gene-encoded peptides was surprising. Biological activity of these peptides depends on the presence of D-amino acids in specific positions. Due to the success of chemical synthesis, it now seems feasible to alter some properties of proteins by incorporating D-isomers of amino acids in certain positions. The inclusion of D-amino acids opens several important scientific and therapeutic implications, including the possibility of protecting peptides from proteolytic cleavage and altering their antigenic properties.

Recently, the group directed by Stephen Kent (Scripps Research Institute, La Jolla, CA) obtained a D-enantiomer of HIV-1 PR by total chemical synthesis and demonstrated that the enzyme had reciprocal chiral specificity toward peptide substrates and inhibitors. We used a large amount of this material, prepared by chemical ligation of backbone-engineered HIV-1 PR, for crystallographic studies.

The D-enzyme and the D-inhibitor in tenfold excess (relative to the PR dimer) were cocrystallized at pH 5.4 (in 20 mM sodium acetate buffer). Crystals grew at room temperature in hanging drops from 50% ammonium sulfate. The protein concentration was 5 mg/ml. The chemical formula of the D-inhibitor used in this work is identical to that of the inhibitor MVT-101 [N-acetyl-Thr-Ile-Nle-(CH_2-NH)-Nle-Gln-Arg.amide], but has an opposite chirality.

We obtained inhibitor-bound [D] HIV-1 PR crystals in the space group $P2_12_12_1$ with unit cell dimensions a = 67.5 Å, b = 92.8 Å, c = 29.4 Å. The crystal morphology is quite different from that of the corresponding L-enantiomer (a = 51.7 Å, b = 59.2 Å, c = 62.5 Å; space group $P2_12_12_1$), which we determined previously, even though the conditions of crystallization are similar in both cases. The crystals diffract to 2.2 Å resolution. However, the present set of data, collected on a Siemens area detector, extends only to a resolution of 2.5 Å. Two platinum derivative data sets, including Bijvoet pairs and extending to a resolution of 2.5 Å and 3.2 Å, respectively, and a mercury derivative data set extending to 3.2 Å were also collected.

In collaboration with J.K. Mohana Rao (Enzyme Structure Section), we have solved the structure of the D-enantiomer by molecular replacement methods. In the absence of anomalous dispersion or other pertinent chemical information, it is not possible to distinguish between L- and D-enantiomers by using X-ray diffraction techniques. The complex of the synthetic HIV-1 PR dimer with the MVT-101 inhibitor, which was refined to an R-factor of 0.154 at 2.0 Å resolution, served as the starting model for molecular replacement. Using the program suite MADIRA developed by Dr. Rao, we obtained the rotation and translation solutions for the PR dimer, followed by a rigid-body refinement. Crystallographic refinement was continued using the X-PLOR and PROFFT programs. At the stage of refinement when the R-factor without the inhibitor was about 0.27 with satisfactory geometry, the hexapeptide was fitted to difference-Fourier electron density maps. We followed the same protocol that was used in the refinement of the L-enzyme so that comparisons between the two structures would be more significant. The present R-factor is 0.188, and further refinement is in progress.

We determined the absolute configuration by using anomalous diffraction data of the platinum derivative. Comparison of the structure with that of the natural-backbone synthetic L-amino acid enzyme showed that the D-enzyme was, in all respects, the mirror image of the L-enzyme (Figure 11). This relationship applied to every aspect of the three-dimensional structure of the homodimeric 22-kilodalton enzyme molecule and extended to centers of asymmetry not directly determined by the chirality of α-carbon atoms in the polypeptide backbone. These data irrefutably show that no chiral effects, other than those inherent to the covalent structure of a polypeptide chain, are required for correct three-dimensional folding to form a protein structure.

FIGURE 11. Stereoview of the crystal structure of the D-enantiomer of the HIV-1 PR showing the difference electron density for the bound D-inhibitor peptide.

Crystal Structures of Rabbit Liver and Muscle Aldolases

In collaboration with Jurgen Sygusch and his colleagues at the University of Montreal, Bangalore K. Sathyanarayana is working on the crystal structure determination and refinement of aldolases from rabbit liver and muscle.

The rabbit liver aldolase crystallizes in space group $C222_1$ with unit cell dimensions a = 377.02 Å, b = 130.35 Å, c = 80.04 Å, diffracting to a resolution of 2.1 Å. The enzyme is a tetramer, with each subunit consisting of 364 amino acid residues. All of the residues in each subunit have been fitted to the electron density maps. The C-terminal residues have been located, which so far has not been reported by any other investigators. In two subunits, the C terminus ends at the active-site region, whereas in the other two subunits, the C terminus ends at the active-site region of their symmetry mates. Also, several substrates have been fit into the active-site region. There seems to be strong evidence for two alternate conformations for the C terminus. The current R-factor of the enzyme is 0.21 without any water molecules. One of the subunits of liver aldolase is shown in Figure 12.

In a manner similar to the above investigation, the crystal structure of rabbit muscle aldolase at 1.9 Å was refined to an R-factor of 0.19. Unlike the liver aldolase, the muscle aldolase crystallizes in a space group of $P2_1$ with unit cell dimensions a = 164.1 Å, b = 57.5 Å, c = 85.10 Å, $\alpha = 90°$, $\beta = 102.6°$, $\gamma = 90°$. However, like the liver aldolase, the muscle aldolase is also a tetramer composed of 364 residues per subunit, and the C-terminal residues of each subunit have been located on the electron density maps for the first time. Both the structures are almost complete, with a total of 800 water molecules located in the case of muscle aldolase.

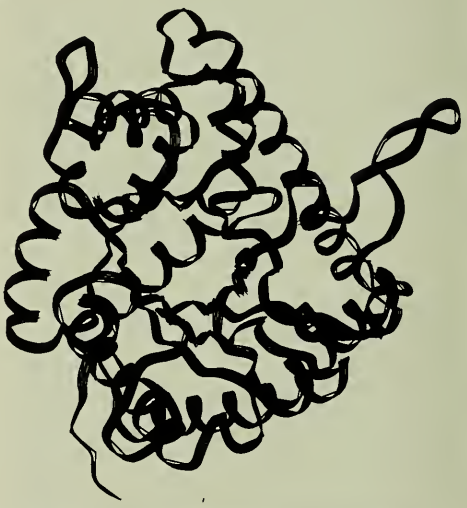


FIGURE 12. The three-dimensional structure of one subunit of rabbit liver aldolase.

Laboratory of Eukaryotic Gene Expression

Jeffrey N. Strathern

Amar J.S. Klar
David J. Garfinkel

The Laboratory of Eukaryotic Gene Expression encompasses three interactive research sections investigating the molecular biology of budding and fission yeast: the Genome Recombination, Rearrangement, and Regulation Section, headed by Jeffrey N. Strathern; the Movable Genetic Elements Section, headed by David J. Garfinkel; and the Developmental Genetics Section, headed by Amar J.S. Klar. These sections share interests in how genomes recombine and evolve as well as in the mechanisms of genetic regulation. Several collaborations span these research sections and a weekly joint research conference helps foster intellectual interaction and the exchange of technical advances.

Research in the Genome Recombination, Rearrangement, and Regulation Section is divided into two areas: the mechanisms controlling gene expression and the mechanisms of genetic recombination. This year, our studies of gene regulation have focused on the role of the transcriptional regulator Ste12p in the cell-type-specific expression of genes involved in mating. Ste12p is required for the expression of a-specific and α-specific genes. We describe several novel *ste12* mutations that cause cell-type-specific defects in gene expression. These mutations may affect the interaction of combinatorial regulatory factors that define cell-type-specific transcription patterns. We have made progress in two major projects related to genetic recombination this year and are also monitoring the accuracy of DNA synthesis associated with recombination. Our results indicate a role for a lower-fidelity DNA polymerase in the repair of double-strand breaks (DSBs). We have identified the polymerase responsible for the majority of these errors and have demonstrated an asymmetry in the strandedness of the errors. Our research on RNA-mediated recombination focused on the priming of reverse transcription of non-retroelement transcripts. Furthermore, we have found that a reverse transcriptase from a LINE-like element from the trypanosome *Crithidia fasiculata* can support the formation of cDNAs in yeast.

Research in the Movable Genetic Elements Section addresses the biology of the yeast retrotransposons Ty1 and Ty2. Members of this section have developed a genetically marked Ty element that allows the selection of Ty retrotransposition and have focused on Ty integration mechanisms.

They used three approaches in studying Ty integration: (1) to distinguish homologous recombination events between Ty cDNA and endogenous Ty elements by their dependence on a gene involved in recombination, *RAD52*; (2) to express Ty integrase (without the rest of Ty) from an inducible promoter and to demonstrate enzymatic activity from the purified Ty integrase; and, (3) to determine the position of several hundred Ty insertions in a target gene and to show that the site specificity of Ty integration is affected by mutations that alter chromatin structure. The second focus of the Movable Genetic Elements Section was the study of how Ty elements inserted into the promoter region of a gene can control the expression of the disrupted gene. In particular, this section has characterized the *SPT23* gene, a locus identified by its influence on the expression of genes that are under the control of nearby Ty elements. These results suggest that Spt23p is a component of chromatin and that loss or overexpression of *SPT23* can alter gene expression.

The Developmental Genetics Section concentrated on the mechanism of silencing of the donor loci (*mat2* and *mat3*) involved in mating-type interconversion in the fission yeast *Schizosaccharomyces pombe*. Members of this section have cloned and sequenced the *clr1* gene (a gene required for silencing *mat2* and *mat3*) and have identified three new genes (*clr2*, *clr3*, *clr4*) that are also involved in this process. Mutations of these genes have several consequences, including allowing the expression of *mat2* and *mat3*, altering the pattern of cell-type switching, and removing a total block on recombination in the *mat2-mat3* interval. This section is presently investigating the roles of the recombination block (or coldspot) on the choice of donor locus and the resolution of the switching intermediate. They have also demonstrated that the *swi2* gene product is required for switches producing cells of altered mating type.

Genome Recombination, Rearrangement, and Regulation Section

Jeffrey N. Strathern

Leslie Derr
Susan Holbeck
Salome La Roche
Carolyn McGill
Dwight Nissley
Brenda Shafer
Keith Weinstock

During the past year, we have addressed four areas of research: control of cell-type-specific gene expression; fidelity of DNA synthesis associated with double-strand break (DSB) repair; initiation of recombination; and mechanism of RNA-mediated recombination. As in the past, our progress is in part a result of our formal and informal interactions with the members of the Movable Genetic Elements Section and the Developmental Genetics Section.

The experiments on cell-type regulation have developed from our long-standing interest in how the different mating types of the yeast *Saccharomyces cerevisiae* are controlled. Last year, we isolated a novel allele of a gene (*STE12*) that is normally a positive regulator of genes involved in mating in both a and α cells (1992 Annual Report). This new allele, *ste12-T50*, allows the expression of a-specific genes, but not α-specific genes. This year, we present further characterization of this mutation and of additional alleles with similar phenotypes.

The experiments related to the fidelity of DNA synthesis during DSB repair represent a continuation of our analysis of mitotic recombination events initiated at defined sites. During the past 2 years, we developed an efficient mechanism of generating interchromosomal recombination events between homologous sequences in mitotically growing cells. We have analyzed recombinators based on a double-strand endonuclease (the HO protein) and an enzyme that cuts one strand of the DNA helix (the gene II protein of bacteriophage f1). Last year, we demonstrated a 100-fold increase in the reversion frequency of mutations (relative to spontaneous mutations) that are located at the site of recombinational repair synthesis. These data were based on a single-base substitution allele (a nonsense mutation). This year, we have analyzed two frameshift alleles (20-base insertions) and have revealed an additional asymmetry in the repair process.

While several attractive models have been proposed for recombination based on single-strand nicks as initiation events, there has been no demonstration that such nicks are sufficient to promote recombination. Our experiments showing that gene II protein could promote recombination in yeast was a step toward such a demonstration. This year, we developed a test to determine whether events initiated by gene II protein could be distinguished from events initiated by HO protein with regard to the fate of flanking sequences.

This year, our analysis of RNA-mediated recombination extends our initial demonstration of both classes of genetic events that have been interpreted as products of this process: precise intron loss, and the formation of processed pseudogenes. Last year, we demonstrated that these events are entirely dependent on the expression of the yeast retrotransposon, Ty. Further, we demonstrated that Ty reverse transcriptase is necessary, but not sufficient, for this process. There is an additional requirement for the replication of the Ty genome that apparently reflects the sequence-specific priming mechanism of Ty reverse transcriptase. We present below our initial characterization of RNA-mediated recombination events promoted by a foreign reverse transcriptase expressed in yeast.

Cell-Type-Specific Gene Expression

The transcriptional activator Ste12p is not only a component of activation complexes that control three distinctly regulated sets of genes, but also the proposed target for the yeast pheromone-responsive pathway. We have isolated and characterized a *ste12* allele with an α-specific mating defect from *matα2* cells. In *matα2* cells, both sets of cell-type-specific genes are expressed simultaneously due to the lack of repression of a-specific genes by the α2 protein. As previously described, *ste12-T50* (which has threonine at codon 50 instead of the wild-type alanine) was found as a second-site mutation that can confer the Alf (a-like faker) phenotype to the *matα2* background. We have continued to investigate the effect of *ste12-T50* on mating and transcription of cell-type-specific genes in isogenic *MAT*a and *MAT*α backgrounds. The differential effect of *ste12-T50* on the mating efficiency of a- versus α-cells is correlated with a differential effect on the expression of a- versus α-specific genes (Figure 1).

***ste12-T50* and Its α-Specific Defect.** *STE12* encodes a transcriptional activator that acts as a key component of the yeast signal transduction system for pheromone response and mating. Ste12p, phosphorylated in response to pheromone, activates three classes of genes that facilitate mating and fusion of the two cell types a and α. Models of the typical upstream activator sequences (UAS) for each of these classes are depicted in Figure 2. The UAS of genes that are pheromone responsive in both a and α cells consists of two or more repeats of the pheromone-response element (PRE). Ste12p binds directly to these sequence elements, and in vitro studies suggest a slight cooperativity in binding of a homodimer of Ste12p to two vicinal PREs. The UAS of a-specific genes has been described as a PRE-box for binding of Ste12p next to a sequence (called a P-box) that binds the general transcriptional activator Mcm1p. The UAS of α-specific genes (called the QP-box) is believed to be recognized by a complex of Ste12p, Mcm1, and the α1 protein. The Q-box is recognized by α1, which binds cooperatively with Mcm1 bound at the P-box. It is not yet known whether Ste12 is in contact with DNA through a weak consensus PRE site near the QP boxes or is bound only through protein-protein contacts to Mcm1 and/or α1.

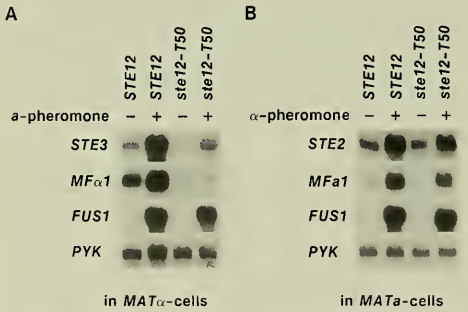


FIGURE 1. Transcription of genes involved in mating. Northern blot analysis of RNA isolated from isogenic *MAT*α (A) and *MAT*a (B) strains that contain either *STE12* or *ste12-T50*. Total RNA was isolated from cells that were not treated with pheromone (–) or from cells that were treated with pheromone for 30 minutes (+). Approximately 20 μg of RNA was loaded per lane, and the *PYK* gene was used as a loading control. The genes used as probes in independent experiments are indicated next to their signal.

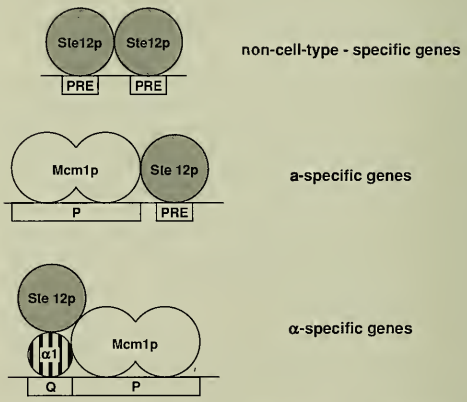


FIGURE 2. Models of the UAS elements of three classes of pheromone-induced genes: non-cell-type-specific, a-specific, and α-specific. PRE, pheromone-response element and binding site for Ste12p (Dolan et al., Proc Natl Acad Sci USA 86:5703, 1989); P, DNA-binding site of Mcm1p (Errede and Ammerer, Genes Dev 3:1349, 1989); Q, DNA-binding site of α1 (Sengupta and Cochran, Genes Dev 5:1924, 1991). It is not clear whether Ste12p contacts the DNA at the QP-box; thus, the depicted model is simplified.

Ste12p is required for both the basal and pheromone-induced levels of transcription. The basal level may reflect a low level of constitutive pheromone response or the activity of the unphosphorylated form of Ste12p. The *ste12-T50* allele was the first *ste12* mutation that selectively affects the expression of α-specific genes. Measurements of the non-cell-type-specific mating gene *FUS1* and the a-specific genes *STE2* and *MFa1* show no effect of *ste12-T50* on the level of expression. In contrast, expression of the α-specific genes *STE3* and *MFα1* is significantly reduced in *MATα ste12-T50* cells (Figure 1).

The most obvious explanation for this α-specific defect is that Ste12-T50p has a reduced ability to interact with α1. Thus, the α1-Mcm1p-Ste12p complex does not form properly to provide α-specific gene activation. An alternative explanation is that the *ste12-T50* mutation abolishes a DNA-binding activity of Ste12p required for formation of the α-specific activation complex, but not for the activation of a-specific genes.

We have recently isolated several additional *ste12* alleles by mutagenesis of the cloned gene and have screened these alleles for mutations that result in phenotypes similar to those caused by *ste12-T50* (Figure 3). Two mutations proved to be independent isolates of *ste12-T50*, and two resulted in valine codons at the same position, *ste12-V50*. As previously noted, this position is within the minimal domain for binding to DNA at the PRE and is just upstream of the region homologous to the homeo domain, a motif common to this class of DNA-binding proteins. A second class of mutations included three independent mutations that altered a highly conserved Arg codon within the first helix of the DNA-binding domain of homeobox proteins. The third class of mutations obtained from this screen were nonsense mutations near the C terminus of the gene. Whether these three classes of mutations reflect the same mechanism is yet to be determined. In addition, we screened a collection of *ste12* alleles generated by the laboratory of Stanley Fields (State University of New York at Stony Brook), and we identified five alleles that resulted in the Alf phenotype in a *matα2* background. Two are in-frame linker insertions near the DNA-binding domain of Ste12p, two are deletions near the C terminus, and one is a short internal deletion. Using this growing collection of *ste12* alleles, we hope to define the domains of Ste12p required for the various contacts involved in the activation of α-specific genes.

FIGURE 3. Ste12p and the mutations that show the Alf phenotype in *matα2* cells. Li-37 and Li-103, in-frame linker insertions (Kirkman et al., Mol Cell Biol 13:3765, 1993); T50, V50, and K61, mutant amino acids and their positions (for wild-type amino acids, see Dolan et al., Proc Natl Acad Sci USA 86:5703, 1989); am374, amber mutation; oc380, ochre mutation; Δ14, Δ7, and Δ22, deletions described by Kirkman et al. (Mol Cell Biol 13:3765, 1993).

DNA Synthesis Errors Associated With DSB Repair

Last year, we described initial experiments addressing the fidelity of DNA synthesis associated with DSB repair. These experiments represent a novel approach to identifying the polymerases involved in DNA synthesis and the mechanisms that determine their error rates. In previously used methods, replication errors were detected as the production of mutants or reversion of defective alleles. Such assays detect the sum of all pathways that can result in DNA alterations. Hence, misincorporation during S-phase replication is not distinguished from events induced directly by DNA damage or events produced by misincorporation during repair synthesis. We have developed an assay that focuses on misincorporations made during a single event — recombinational repair of a site-specific DSB.

Last year, we demonstrated that the reversion rate for a nonsense allele is 100-fold higher in cells undergoing repair of a DSB introduced 300 bases from the mutation than the spontaneous reversion rate for the same allele. This year, we investigated the reversion rate of frameshift alleles and again observed that it was over 100-fold higher than spontaneous reversion rates in the cells repairing a DSB near the frameshift allele. Further, we observed that the nature of the reversion events is different for the spontaneous and DSB repair-associated events.

The design of these experiments involves the use of HO endonuclease to create a unique DSB on one chromosome in a diploid and monitor the fidelity of the repair synthesis associated with repair of that break. In yeast, DSB repair is almost exclusively accomplished by recombination in which a homologous sequence is used as a template for DNA synthesis spanning the site of the break. For our experiments, we placed the HO cleavage site within a region specifically constructed for recombination studies; this region includes the *trp1* and *his3* genes inserted on chromosome III near *MAT*. The same *trp1* allele was present on both chromosomes in these diploids (Figure 4A). Two different frameshift alleles were generated by the insertion of a 20-base pair oligonucleotide into the *Hin*dIII site within the *trp1* open reading frame (Figure 4B). The HO site was present on only one chromosome between the *trp1* and *his3* genes; therefore, recombinational repair of the DSB caused by HO cleavage results in the loss of the HO site.

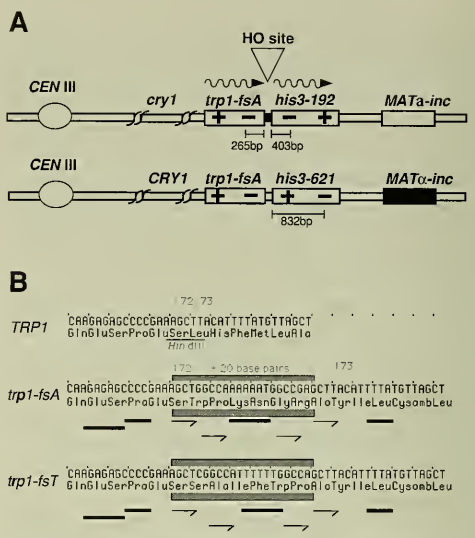


FIGURE 4. Assay for DSB repair errors. (A) Locus for detecting reversion of frameshift mutations. The *trp1* and *his3* genes are inserted on chromosome *III* centromere proximal to *MAT*. Both chromosomes have the same *trp1* allele, a frameshift caused by the insertion of 20 bases at a *Hin*dIII site. The two chromosomes carry different *his3* mutations. A recognition site for the HO endonuclease (HO site) was inserted into a polylinker between *trp1* and *his3* on one chromosome. The normal recognition sites for HO at *MAT* are mutated by the *MAT*a-*inc* and *MAT*α-*inc* alleles. (B) *trp1* frameshift alleles. A 20-base oligonucleotide was inserted at the *Hin*dIII site in *TRP1*. In one orientation of the oligonucleotide (*trp1-fsA*), there are six tandem adenines in the coding strand, whereas the other orientation (*trp1-fsT*) has six tandem thymines in the coding strand. Deletion between the pairs of arrows indicating short direct repeats restores Trp^+. The bars indicate repetitive nucleotides at which deletion of two bases or addition of one base restores Trp^+.

Each cell can undergo only one HO-induced recombination event. We monitored reversion of a *trp1* allele adjacent to the HO site under HO-induced and -repressed conditions. The normal sites for HO recognition in the *MAT* alleles were mutated so that they would not be substrates (*MAT*a-*inc* and *MAT*α-*inc*). The presence of two different alleles of *his3* provided a means to monitor HO-induced recombination. We observed the efficient production of *HIS3*

recombinants (up to 3% of cells after HO induction) in strains carrying this configuration of the HO site and the *his3* alleles.

***TRP1* Reversion Was Elevated Among the HO-Induced Recombinants.** Yeast strains with the configuration of alleles shown in Figure 4A were transformed with the pGALHO plasmid or a related plasmid, pGAL*lacZ*, to make control strains. Reversion of the *trp1-fsA* allele was measured for cultures grown in the noninducing carbon source raffinose and, again, 12 hours after shifting the cultures to galactose-containing medium. The results, shown in Table 1, indicate a 30-fold increase in the frequency of Trp^+ cells for the pGALHO strain. At this time point, about 9% of the cells had experienced an HO-induced recombination event as determined by loss of the HO site or a change in the configuration of other alleles in the *cry1 trp1-his3 MAT* interval. Combined, these numbers suggest a 300-fold higher reversion frequency among the cells that had undergone this DSB repair. Similar to the results observed previously for the reversion of a nonsense allele, the reverted allele was preferentially inherited by the chromosome that experienced the HO cleavage. No increase in the frequency of Trp^+ cells was evident for the pGAL*lacZ* control strain, and the revertant allele was found equally on both chromosomes.

We made the two frameshift alleles by inserting a 20-base oligonucleotide into the *Hin*dIII site in *TRP1*. This region of the Trp1 protein is very tolerant of alterations. We have identified 18 different kinds of revertant alleles of mutants in this interval. In one orientation, the oligonucleotide (called *trp1-fsA*) introduces six adenines in the coding strand. We expected that this region would present a favorable place for polymerase slippage. Single-base insertions in this interval generated a *Bst*XI site and represented the majority event among HO-induced revertants (12/14) and spontaneous revertants (6/9). For the oligonucleotide in the other orientation (six thymines in the coding strand; designated *trp1-fsT*), only 3/27 revertant alleles were the result of insertions of an extra T in this interval. The majority (17/21 sequenced) of the HO-induced revertants of the *trp1-fsT* allele were the result of single-base insertion at a specific position, which created a palindrome. This class of revertants comprised only 2/13 of the spontaneous revertants of *trp1-fsT*. The insertion that made the corresponding palindrome was found in only 1/14 HO-induced revertants of *trp1-fsA* and was not found among the spontaneous revertants of that allele. The different spectra of HO-induced revertants of *trp1-fsA* and *trp1-fsT* suggest that the two strands of DNA at that position are treated differently in the DSB repair process. The simple copying of one strand through the *trp1-fsA* allele should be the same as copying the other strand through the *trp1-fsT* allele. The observed asymmetry is consistent with the proposal that most of the repair synthesis that spans the mutation site is confined to a specific strand. At a DSB, there appears to be more extensive degradation of the 5′-end strand than of the 3′-end strand, which would require a similar asymmetry in the repair synthesis. We suggest that DNA synthesis primed by the invasion of the uncut duplex by a 3′ end on the *his3* side of the DSB crosses into the *trp1* gene. This new strand, containing any errors made during its synthesis, could be inherited by the chromosome that suffered the DSB by pairing with the strand that ends 3′ on the *trp1* side of the DSB (Figure 5). This model not only explains the observed asymmetry in the inheritance of the errors on the cut versus the uncut chromosome, but also is consistent with the different ratio of the types of revertants from the *trp1-fsT* and *trp1-fsA* alleles.

TABLE 1. Reversion of *trp1* frameshift mutations

Mutation	Selection medium	Frequency at mutations in pGAL HO	
		$Trp^+ \times 10^9$	$His^+ \times 10^5$
trp1-fsA	Raffinose	2.5	1.7
	Galactose	86*	610*
trp1-fsT	Raffinose	1.0	1.5
	Galactose	186†	756†

*At this point, only 9% of the cells have lost the HO site.
†At this point, only 3% of the cells have lost the HO site.

The *REV3* Polymerase Causes the Majority of the Errors During DSB Repair. Some types of chemical- and radiation-induced DNA damage are repaired by a process that involves replication across the damaged base(s). This translesion repair synthesis mechanism is referred to as error-prone repair. In yeast, this synthesis is believed to be performed by Rev3p, the product of a gene (*REV3*) with homology to DNA polymerases and required for mutagenesis induced by ultraviolet light. The DNA polymerases responsible for

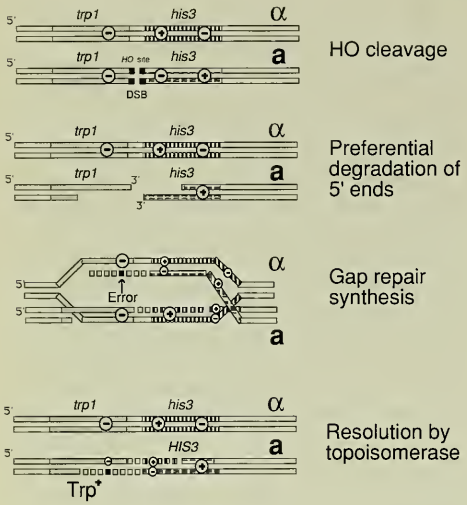


FIGURE 5. Model for asymmetric DNA synthesis and, hence, asymmetry in the strandedness of replication errors during repair of DSBs. Stippled regions represent new DNA synthesis. A replication error is shown in black.

S-phase replication are not believed to be involved in translesion repair processes. In contrast, DNA damage that results in DSBs is repaired by recombinational repair and involves a different set of genes than does error-prone repair. Most models of recombinational repair invoke exonucleolytic degradation of the damaged DNA to expose recombinogenic single-strand tails, with DNA synthesis copying an undamaged template to restore the lost sequences. We have identified *REV3* as one of the genes used in DSB repair.

We introduced a *rev3::LEU2* knockout allele into our strains to determine whether Rev3p has a role in the production of mutations at the site of DSB repair. *REV3* has been reported to have no major role in the formation of recombinants. Consistent with this conclusion, we observed that the induction of HO endonuclease in the *rev3* tester strain resulted in no loss of viability and a several hundredfold increase in the formation of *HIS3* recombinants. However, there was only a twofold increase in the reversion rate of the *trp1-488* allele as compared to the 100-fold increase in *REV3* strains. Thus, Rev3p is responsible for the majority of the replication errors detected at the site of DSB repair. The fact that *rev3* strains retain recombination proficiency implies that another polymerase is also involved in DSB repair. We are now conducting a mutant search to identify other genes that affect the fidelity of DNA synthesis near a DSB.

Summary. DNA damage repair is usually classified as error-prone if it involves synthesis across a DNA lesion or error-free if it involves removal of the lesion and copying of an undamaged template. Recombinational repair is considered to be a subset of the error-free repair processes. The experiments described here monitored the reversion of a mutation located nearly 300 bases from the site of an inducible DSB. It is presumed that cleavage by HO causes no direct damage at the site of the *trp1* alleles and that the observed elevated frequency of revertants among the HO-induced recombinants is the result of errors made during the DNA synthesis associated with the DSB repair process. These observations provide us with a novel opportunity to study the mechanisms that govern the fidelity of a specific DNA replication event. We are developing screens for mutants that give higher (or lower) error rates associated with this repair synthesis, with the hope of genetically defining the components of this process.

Initiation of Recombination by Nicks May Involve Their Conversion to DSBs

We have demonstrated that the gene II protein of bacteriophage f1, an enzyme that makes a site-specific single-strand nick, is able to stimulate interchromosomal mitotic recombination in the vicinity of its recognition site. The level of stimulation induced by gene II protein was much lower than the level induced by DSBs introduced in the same interval by the HO endonuclease, but it is not clear whether this decrease reflects a difference in the inherent recombinogenic nature of nicks versus DSBs or a difference in the activity of gene II protein versus HO in finding and cleaving their respective sites. As in HO-induced recombination events, the chromosome with the recognition site for gene II protein acted as the recipient in gene conversion events. Unlike the expected result for a DSB, gene II protein stimulated recombination preferentially on one side of the site. While this latter result is consistent

with the proposal that these events are nick initiated, it does not constitute proof that the gene II protein-generated nicks do not become DSBs. In particular, in the presence of Mn^{2+}, gene II protein is known to form DSBs in which one side of the break is a hairpin. This asymmetry could account for the directional bias observed in the stimulation of gene conversion.

To address whether gene II protein-initiated events occur through a DSB intermediate, we developed an assay to distinguish nick-mediated versus DSB-mediated events. The principle of the assay is that the region containing the site for a DSB-generating enzyme will be lost during recombinational repair with the template (uncut) chromosome, which does not contain the site. In contrast, events generated by nicks could propagate outside of the region of nonhomology between the nicked chromosome and the template without losing the recognition site for gene II protein. In our experiments, we included the *TRP1* gene within the region of nonhomology between the chromosome with the enzyme recognition site and the template chromosome (Figure 6). We found that both HO and gene II proteins could stimulate the production of *HIS3* recombinants in strains containing their respective sites. As expected, nearly all (524/526) of the His^+ recombinants that were generated when HO was induced resulted in the loss of the *TRP1* gene. This frequency suggests that the DNA nonhomologous to the template chromosome (including the *TRP1* gene) must be removed before the DSB can be repaired. In contrast, most of the spontaneous His^+ recombinants retained the *TRP1* gene. For the His^+ recombinants generated when gene II protein was induced, 95% of the recombinants lost the *TRP1* gene. The important issue is whether any of the events that retained the *TRP1* gene (and the gene II protein recognition site) reflect events that were initiated by a gene II protein-generated nick that did not subsequently become a DSB. There is no evidence to support this conclusion. The few Trp^+ His^+ recombinants obtained were present at about the frequency predicted for spontaneous recombinants among the gene II protein-induced events. These data suggest that gene II protein-induced recombinants result from a mechanism that is not distinguishable in this assay from DSB-initiated (HO-induced) events.

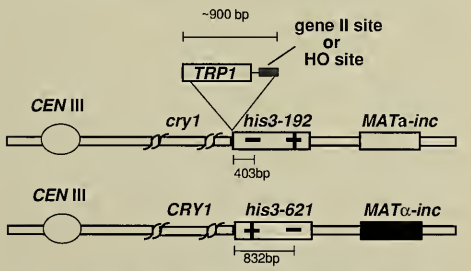


FIGURE 6. Diploid used to monitor the fate of the gene II site or the HO site in *HIS3* recombinants stimulated by these enzymes. The recognition sites for these enzymes are included with *TRP1* as a 900-base insertion on one chromosome.

RNA-Mediated Recombination

Last year, we demonstrated RNA-mediated recombination events in yeast that mimicked two natural processes: the loss of introns from genes and the formation of processed pseudogenes. These processes occur over evolutionary time scales and have not previously been amenable to analysis. We demonstrated that these recombination events are entirely dependent on the expression of the yeast retrotransposon, *Ty*. We suggested that the pathway for the production of these RNA-mediated recombination events involves the production of a cDNA intermediate. We demonstrated that the reverse transcriptase of *Ty* is necessary, but not sufficient, for this process. An additional requirement for replication of the *Ty* genome apparently reflects the sequence-specific priming mechanism of *Ty* reverse transcriptase.

Our assay uses the plasmid-based *GAL1::his3-AI* indicator gene, which has an artificial intron inserted into *his3* in the orientation that allows it to be spliced from the antisense transcript (expressed from the *GAL1* promoter), but not from the sense transcript. Cells with the *his3-AI* allele are His^-. His^+ cells are generated in this system by RNA-mediated recombination: reverse transcription of the spliced *GAL1::his3-AI* transcript followed by recombination. We have distinguished two classes of such $HIS3^+$ recombinants: those that involve the insertion of the *HIS3* gene into a novel location, and those that involve homologous recombination with *his3* (either by loss of the intron from the plasmid generating *GAL1::HIS3* or by the gene conversion of a chromosomal *his3* allele to *HIS3*).

The *HIS3* sequences inserted at novel sites are processed pseudogenes: they are complete copies of the spliced *GAL1::his3-AI* transcript, including the correct initiation site and a polyA tail. These pseudogenes are flanked by *Ty* sequences, consistent with a role for *Ty* in their priming and insertion. We found that a *Ty* element that could provide reverse transcriptase, but could not itself replicate due to alterations in its tRNA primer binding site, could not support the formation of *HIS3*$^+$ prototrophs. The absence of both the *Ty*-flanked pseudogenes and the *HIS3* events generated by homologous recombination suggests that all the *GAL1::HIS3* cDNA is made by a template switch, with replication initiating on the *Ty* transcript by its normal priming mechanism and extending onto the spliced *GAL1::HIS3* transcript. Our recent goal has been to provide alternative (*Ty*-priming-independent) mechanisms for the formation of cDNAs and, hence, RNA-mediated recombination events.

The Reverse Transcriptase of a LINE-Like Element Can Support RNA-Mediated Recombination. Our first attempt to develop a *Ty*-priming-independent system for RNA-mediated recombination used the CRE reverse transcriptase from the trypanosome *Crithidia fasiculata*. *CRE* is a LINE-like retrotransposon lacking the directly repeated sequences, termed long terminal repeats (LTRs), that are present in *Ty* elements and retroviruses. The mechanism of replication and integration of LINE elements is not clearly understood, but because they lack LTRs, these processes must be different than those for retroviruses or *Ty* elements. Any model of LINE element replication and integration must incorporate the maintenance of the polyA tail. For example, the reverse transcriptase of LINE-like elements such as CRE may be capable of priming directly on polyA tails.

In our experiments, the CRE reverse transcriptase was expressed from the *GAL1* promoter as a fusion to *Ty* sequences, replacing the *Ty* integrase, reverse transcriptase/RNaseH, and 3′ LTR (*GAL1-TyA-CRE*, generously provided by Abram Gabriel). We used this plasmid as the source of reverse transcriptase to measure RNA-mediated recombination (formation of *HIS3*$^+$ from the *GAL1::his3-AI* transcript). We found that *HIS3*$^+$ prototrophs were formed even when expression of endogenous *Ty* elements was blocked by the *spt3* mutation (Table 2; *spt3-101*). To further demonstrate that the *HIS3*$^+$ recombinants were independent of the normal *Ty* priming mechanism, we destroyed the complementarity between the primer-binding site of the *Ty/CRE* fusion and the tRNA used by *Ty* reverse transcriptase as primer. Again, *HIS3*$^+$ prototrophs were still produced (Table 2; *TyA-EIB-CRE*). We are presently analyzing the physical structure of the chromosomal *HIS3*$^+$ insertions generated in this system to determine whether the priming and insertion are indeed different than what we previously observed. By using reverse transcriptases from different retroelements and further modifying the template, we may be able to free the system from its dependence on *Ty* biology and increase our understanding of RNA-mediated recombination.

TABLE 2. RNA-mediated recombination with CRE reverse transcriptase

Strain	Plasmid	His3$^+$ x 10^9
SPT3	*GAL1-TyA-CRE*	239
spt3-101	*GAL1-TyA-CRE*	33
SPT3	*GAL1-TyA-EIB-CRE*	478
spt3-101	*GAL1-TyA-EIB-CRE*	39

Developmental Genetics Section

Amar J.S. Klar

Michael Bonaduce
Amikam Cohen
Shiv Inder Grewal
Jagmohan Singh
Genevieve Thon

The fission yeast *Schizosaccharomyces pombe* switches mating type by transposing a copy of the unexpressed *mat2* or *mat3* locus to the transcriptionally active mating-type locus (*mat1*) through a gene-conversion event (Figure 1). The recombination event is initiated by a double-strand break (DSB) found at *mat1*. The pattern of *mat1* switching in cell pedigrees is nonrandom. Earlier, we had proposed a strand-segregation model to explain the pattern of switching in cell pedigrees. Two interesting features of the mating-type system are unusual in that the *mat2* and *mat3* loci are unexpressed and recombination in the 15-kilobase sequence located between *mat2* and *mat3* is totally prohibited. The trans-acting genes *clr1*, *swi6*, and *rik1* are known to silence the *mat2* and *mat3* donor loci and prohibit recombination in the *mat2-mat3* intervening interval.

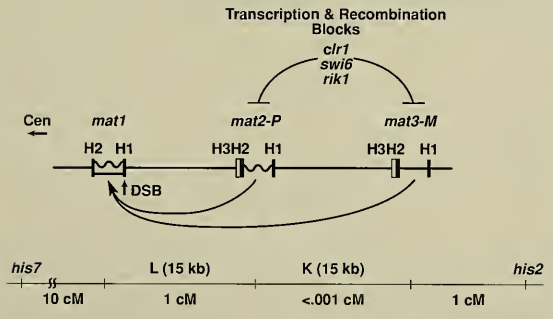


FIGURE 1. Mating-type region on chromosome II. DSB marks the site of the double-strand break. H1 (59 base pairs) and H2 (135 base pairs) are homology boxes flanking the allele-specific 1.1-kilobase cassette sequences.

In the past year, we have (1) characterized the *clr1* gene and identified additional genes that are required to silence the donor loci; (2) obtained evidence suggesting that a site-specific recombination function initiates switching such that a protein "linker" at the cut site exists; and (3) determined that the *swi2*$^-$ gene function regulates the donor choice during switching.

The Mechanism of *mat2-mat3* Silencing

Promoters of transcription, identical to the ones that are present and active at *mat1*, are inactive at *mat2* and *mat3*. The *clr1* locus (*c*ryptic *l*oci *r*egulator) is involved in this repression. During the course of our studies of the mating-type region and of *clr1*$^-$ mutants, we discovered two additional features of the repression mechanism. First, repression is not promoter specific, as other genes (such as the *S. pombe ura4* gene or the *Saccharomyces cerevisiae LEU2* gene) are also repressed when introduced in the vicinity of the silent loci. Second, repression of transcription is associated with repression of the block to meiotic recombination in the *mat2-mat3* interval. Our observation that mutations at *clr1* derepress both transcription and recombination indicates that these processes are blocked by the same mechanism. These features suggest that silencing is promoted by a specific chromatin structure and that our studies will reveal how the expression of linked gene domains can be coregulated by long-range controls.

In the past year, we have obtained genomic clones and determined the DNA sequence of *clr1*. We have also identified three additional linkage groups that are involved in the transcriptional silencing of the mating-type region: *clr2*, *clr3*, and *clr4*.

Phenotypes Conferred by Mutations at the *clr1* Locus. The *clr1*$^-$ mutation confers several phenotypes that can be assayed and used for selections. First, coexpression of both mating types in haploid cells triggers meiosis and an aberrant sporulation phenotype. This "haploid meiosis" phenotype can be observed either microscopically or at the colony level by iodine staining of the spores. Second, auxotrophy markers placed in the vicinity of *mat2* or *mat3* are repressed in wild-type cells and derepressed in *clr1*$^-$ mutants. In several experiments, we used an *S. pombe ura4* gene introduced 150 base pairs distal to *mat3*. In wild-type strains, this gene is repressed, which allows growth on the toxigenic substrate 5-fluoroorotic acid (FOA) but not in the absence of uracil. In the presence of mutations at the *clr1* locus, the *ura4* insert is derepressed, allowing growth in the absence of uracil. In addition, expressed *ura4* converts FOA to a toxic compound, as indicated by lack of growth on medium with FOA (Figure 2).

Cloning the *clr1* Gene. By complementing the *clr1-5* mutation, we obtained three identical genomic clones of 8.5 kilobases containing the *clr1* gene. These clones were selected for their ability to restore repression of the *ura4* gene near *mat3* in *clr1-5* cells grown on medium with FOA. We used one of the clones for further studies. Its insert integrated by homologous recombination at the *clr1* locus, confirming its identity as the *clr1* gene. To localize the *clr1* gene within the clone, we subcloned the insert and identified a 3.5-kilobase *Hin*dIII restriction fragment that rescued the *clr1-5* mutations by homologous recombination. We mutagenized this 3.5-kilobase fragment with the bacterial Tn10 transposon. When introduced in the chromosome of a wild-type strain, the Tn10 insertions shown in Figure 3 conferred a *clr1*$^-$ phenotype, indicating disruption of the gene.

The *clr1* Gene Encodes a Zinc-Finger Protein. Having localized the *clr1* gene within the genomic clone, we then determined its DNA sequence. The *clr1* gene encodes a putative protein of 1238 amino acids. To date, the most significant feature of the predicted Clr1 protein is the presence of three zinc-finger motifs at its C terminus. The presence of these motifs suggests that the protein may bind DNA. The three zinc-finger motifs of Clr1 have sequences

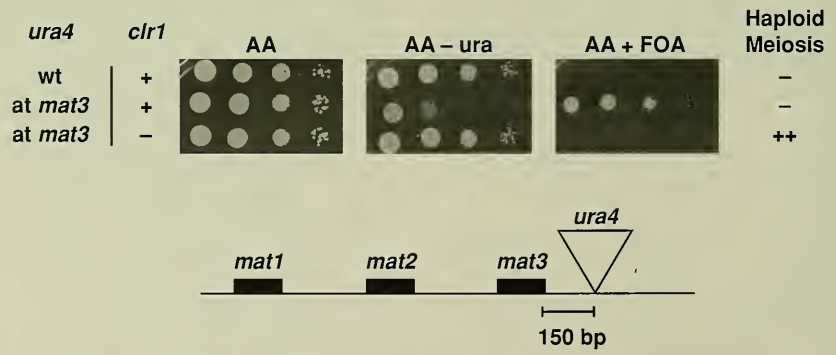


FIGURE 2. *clr1* affects gene expression in the mating-type region. Expression of the *S. pombe ura4* gene placed in the vicinity of *mat3* and expression of the mating-type genes *mat2* and *mat3* were assayed in *clr1*$^+$ and *clr1*$^-$ cells. Serial tenfold dilutions of cell suspensions were spotted on complete medium (AA), medium lacking uracil (AA-ura), or medium containing FOA (AA+FOA). Cells in the first row express *ura4* from its wild-type chromosomal location; cells in the second row have their sole copy of *ura4* near *mat3* and express *ura4* poorly; cells in the third row also have their sole copy of *ura4* near *mat3* and express *ura4* well because they contain a mutation at the *clr1* locus. "Haploid meiosis," a phenotype caused by the expression of *mat2* or *mat3*, was assayed by microscopic examination of sporulated cultures.

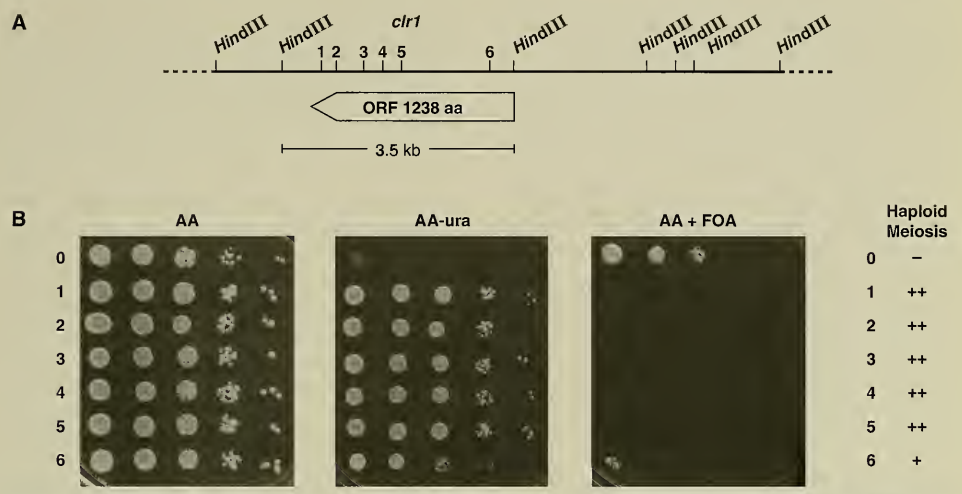


FIGURE 3. Cloning and sequence analysis of *clr1*. (A) Partial restriction map of the 8.5-kilobase genomic clone that complements the *clr1-5* mutation. The numbers 1 to 6 indicate the position of Tn10 insertions obtained within the insert. A long open reading frame (ORF) predicted from the sequence analysis is represented under the restriction map. (B) Mutagenesis of the complementing clone. The three panels show the effect of the Tn10 insertions integrated in the *clr1* gene on the expression of the *ura4* gene placed near *mat3*. Serial tenfold dilutions of cells containing *ura4* near *mat3* were spotted on different solid media. The haploid meiosis sporulation phenotypes that reflect expression of *mat2* and *mat3* are also indicated. Numbers 1 to 6 represent the strains with Tn10 inserts depicted in panel A, whereas 0 indicates the wild-type *clr1*$^+$ strain

similar to three other zinc-finger proteins: the product of the *Caenorhabditis elegans* sex-determining gene *tra-1*; the family of human Gli proteins (so named because one of the *GLI* genes was amplified in a glioblastoma); and the *Drosophila melanogaster* segment polarity gene *cubitus interruptus Dominant*. Because Clr1 does not display significant sequence similarity with these three proteins outside the putative zinc fingers, it seems unlikely that these genes are functional homologs.

We are currently expressing domains of Clr1 in *Escherichia coli* to raise antibodies against the protein, to determine whether Clr1 binds DNA, and to determine the sequence of its binding site. We are particularly interested in determining whether Clr1 binds DNA within the mating-type region, which would indicate a direct involvement in silencing. We are also searching for binding sites in other regions of the genome, particularly where transcription and/or recombination are known to be repressed, such as centromeric and telomeric regions.

Identification of Additional Genes Involved in Silencing. No systematic search for mutants deficient in silencing had been performed in *S. pombe* until now. To identify more of the genes involved in the transcriptional repression of the mating-type region, we mutagenized 20 independent cultures of cells containing the *ura4* gene near *mat3* and selected derepressed mutants able to grow on medium lacking uracil. We also examined the sporulation phenotypes of these mutants. Based on our genetic analysis of approximately 200 mutants expressing the *ura4* gene and one or both of the mating-type genes *mat2* and *mat3*, we defined three linkage groups, designated *clr2*, *clr3*, and *clr4*. We are currently assaying whether mutations at these loci affect *mat2* and *mat3* differently and affect recombination in the *mat2-mat3* interval. With the combined results of this assay and a similar search that we performed using an unexpressed *ura4* gene inserted near *mat2*, we will be able to classify the mutations that derepress *ura4* at *mat2* and compare them with the mutations that derepress *ura4* at *mat3*.

A Mutation Suggesting a Link Between Mating-Type Switching and Reestablishment of Silencing

A mutation in *S. pombe* in which the silent cassettes are turned on was serendipitously found. This mutation, termed *hm1-1*, confers the haploid meiosis phenotype to the cells. Interestingly, the loss of silencing is evident only in strains that are actively switching. The haploid meiosis phenotype is not displayed by strains that do not have the DSB at *mat1* and therefore do not switch (such as *swi7-1*, *mat1Msmt0*, or *mat1PΔ17*) or by strains that have a reduced efficiency of switching (such as *swi2* and *swi5*). This observation suggests that the loss of repression requires switching. Previous studies of the *S. cerevisiae* mating-type system have indicated a link between repression and the utilization of a replication origin located near the silent locus. Thus, in the *hm1-1* mutant, the chromatin structure at the donor locus may not be reset into the repressed state after switching. Interestingly, this mutant is temperature sensitive for growth at 37°C, indicating that *hm1-1* encodes an essential gene. We have genetically mapped this mutation within 10 centimorgans of the *hta1-htb1* locus, which encodes the histone H2A.1 and H2B.1 genes, and we are presently testing whether *hm1-1* would in fact encode one of these histone genes. If it does, then this result would suggest that a repressed chromatin structure is reestablished following donor utilization during mating-type switching.

A Site-Specific Recombination System Is Suggested to Initiate Mating-Type Switching

Two major fascinating aspects of the mating-type switching system are that only one among four granddaughters of a cell ever switches and the sister of the recently switched cell usually produces a switched daughter and therefore is competent to switch. Our past work has established a strand-segregation model that explains the pattern of switching in cell pedigrees. An important feature of the model is that a DSB at *mat1* is made in one cell cycle and repaired in the following cycle when one of the daughter cells switches. This feature creates a major problem in maintaining the integrity of the broken chromosome throughout the cell cycle. Based on the genetic evidence presented below, we suggest that the break is made by a site-specific recombination mechanism, similar to that of prokaryotic Hin and Gin invertases and Tn3 γ δ resolvase.

It is known that a hotspot for mitotic recombination exists at *mat1*. The hotspot is defined as the region where diploid cells undergo a high rate of mitotic crossing-over (i.e., in 4% of the cell divisions) at the mating-type locus. If one uses a diploid strain that is heterozygous for a marker such as *his2*, which is located distal to *mat1*, in 2% of the cell divisions one daughter becomes $his2^-/his2^-$ and the other daughter becomes $his2^+/his2^+$ (Figure 4). All other markers that are located distal to *mat1* also undergo cohomozygosis. In the other 2% of the cell divisions, the coupling of *mat1*–flanking markers is changed (Figure 4). Such a result is explained by recombination occurring at *mat1* in the G_2 phase of the cell cycle, when two chromatids are involved in crossing-over. Based on an assortment of centromeres, recombination is expected to cause homozygosis in half of the cell divisions when the chromatids have crossed over, while in the other half the linkage of *mat-1*–flanking markers will change only in one of the daughter cells (Figure 4). The high rate of crossing-over is thought to be initiated by the DSB at *mat1*.

The explanation of crossing-over occurring by a general recombination mechanism is inconsistent with the following finding. If the diploid contains one wild-type *mat1* allele

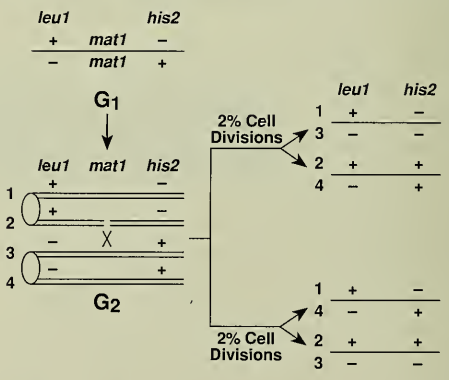


FIGURE 4. A hotspot of mitotic crossing-over at *mat1* in diploid cells. A total of 4% of the cell divisions produce crossing-over at *mat1*. One-half of these divisions lead to homozygosis of markers (e.g., *his2*) that are situated centromere distal to *mat1*. The hotspot of recombination is thought to be initiated by the DSB at *mat1* during the G_2 phase of the cell cycle. Numbers 1 to 4 represent an assortment of specific chromatids in mitosis.

that can acquire the DSB whereas the other allele cannot because it contains a *mat1* cis-acting deletion prohibiting generation of the DSB, then homozygosis does not occur. As the rate of recombination is limited by the efficiency of initiation of recombination, this diploid should be able to undergo homozygosis by general recombination in 1% of the cell divisions as opposed to the 2% rate found in standard diploid strains. Thus, the paradox is that homozygosis occurs when both *mat1* loci are wild type (and therefore can be cut in vivo) but does not occur when one of the *mat1* loci is not cleavable.

We have sought to resolve this paradox by considering site-specific recombination systems, such as those of prokaryotic Gin and Hin invertases and Tn3 γ δ resolvase, which recombine two DNA duplexes by a concerted cleavage and rejoining reaction. Site-specific recombination reactions bring about controlled rearrangements of DNA molecules by cutting the DNA at precise points and rejoining the ends to new partners with no synthesis or degradation of DNA. This mechanism involves a covalent intermediate between a serine hydroxyl residue on the enzyme and 5′-phosphate in the DNA backbone. This bond conserves the high energy of the phosphodiester bond, so that subsequent ligation to 3′ hydroxyls in the DNA can proceed without exogenous energy carriers such as ATP. The last step in the reaction is the reversal of cleavage. Four subunits of the enzyme participate in the reaction, with two subunits bound at each crossover site. The four DNA ends and the proteins covalently attached to each of them form a symmetrical complex that is held together by interactions between the proteins. One 180° rotation of two subunits with respect to the other two and religation of the swapped DNA chains complete the reaction. This mechanism requires that both duplexes be cleaved simultaneously in order for a swapping reaction to occur, which is reminiscent of the requirement of having both wild-type *mat1* loci in the diploid for homozygosis. Thus, it is possible that homozygosis does not occur by normal recombination; rather, homozygosis may occur by a site-specific recombination ("swapping") mechanism.

Three observations are consistent with this model. First, based on the level of the DSB found in vivo, the observed level of crossing-over is predicted by the site-specific recombination mechanism. Second, the model suggests that recombination occurs by swapping cleaved DNA duplexes and not by the usual double-stranded break repair mechanism. Therefore, we predict that recombination at the hotspot should also occur in recombination-deficient mutant strains. In agreement with this prediction, we found that homozygosis is not reduced in *swi5* mutants. Furthermore, *swi5*$^-$ mutants are known to be deficient in general recombination as well as in mating-type switching that is promoted by the DSB at *mat1*. The lack of repair/recombination is also suggested, as *swi5* mutants are ultraviolet- and γ-ray–sensitive. Third, our previous results have shown that both 5′ ends of the cut DNA are resistant to in vitro digestion by the 5′-specific lambda endonuclease. Therefore, proteins that cleave *mat1* in vivo may hold together the cleaved ends and thus maintain the integrity of the chromosome throughout the cell cycle. In other words, we propose that a protein "linker" holds the cut ends together. Such proteins at the cut may promote interaction with the donor loci for the strand-transfer reaction. It is noteworthy that the homology sequences H1 (59 bp) and H2 (135 bp), which are required for gene conversion for *mat1* switching, are very short, yet *mat1* switching occurs efficiently. The proposed site-specific recombination proteins may promote synapsis by protein-protein and protein-DNA interactions instead of by extended DNA homology. We will direct our future experiments toward identifying the suggested cleaving enzyme. These results suggest that the mechanism of initiation of mating-type switching is fundamentally different in *S. pombe* and evolutionarily distantly related to that of *S. cerevisiae* budding yeast.

The *swi2* Gene Regulates Donor Choice During Switching

Because switching-competent cells switch to the opposite allele in 80% to 90% of the cell divisions, cells must preferentially choose the donor locus that contains information different from that existing at *mat1*. In wild-type strains, *mat2* and *mat3* contain the *P* and *M* alleles, respectively. Thus, *P* cells choose *mat3* preferentially, whereas *M* cells choose *mat2*. Our past work with a strain whose donor content at *mat2* and *mat3* was swapped demonstrated that donor loci are selected for their genomic location and not for their information content (1992 Annual Report).

We are interested in determining how the donor locus is selected. One possible candidate for such a function is the *swi2* gene. The *swi2*$^-$ mutants have a normal level of the DSB at *mat1*, yet cells switch inefficiently to the opposite allele.

One possibility is that the Swi2 protein may be needed to further process the DSB for recombination. Meiotic *mat1* conversion, also initiated by the same DSB, is not affected, suggesting that Swi2 is not directly required to process the DSB for gene conversion.

Another possibility is that Swi2 function may be required to promote switching only to the opposite direction. That is, *swi2*$^-$ mutants may undergo futile homologous (*M* to *M* and *P* to *P*) switches. It was previously known that a strain containing a mutation in *mat2-P* (*mat1-P mat2-P mat3-M*) efficiently replaces the wild-type *mat1-P*$^+$ allele by switching, suggesting that *swi2*$^-$ strains switch efficiently. Yet, *swi2*$^-$ strains containing wild-type *mat2-P* and *mat3-M* cassettes switch inefficiently to the opposite allele. One solution for this paradox is that *swi2*$^-$ mutants preferentially undergo homologous cassette replacement. To test this possibility, we determined the effect of donor loci genetic content on the efficiency of switching to the opposite allele in *swi2*$^-$ mutants. We found that strains with the genotypes *swi2*$^-$ *mat1-P mat2-M mat3-P*, *swi2*$^-$ *mat1-M mat2-M mat3-P*, *swi2*$^-$ *mat1-P mat2-M mat3-M*, and *swi2*$^-$ *mat1-M mat2-P mat3-P* switched very inefficiently to the opposite allele as compared to the *swi*$^+$ wild-type control strains (Table 1). In other words, *swi2*$^-$ mutants are able to switch efficiently when both *P* and *M* donor information is present in the strain, but switching is greatly reduced when both donors contain cassette information opposite to that found at *mat1*. Thus, we conclude that Swi2 function is required for switching only to the opposite allele.

TABLE 1. Effect of *swi2* on directionality of switching

Cassettes					
mat1	*mat2*	*mat3*	*swi2*	Direction	Efficiency
P, M	*P*	*M*	+	*P*↔*M*	High
P, M	*P*	*M*	–	*P*↔*M*	Low
P, M$^-$	*P*	*M*$^-$	–	Unknown	High
P$^-$, *M*	*P*$^-$	*M*	–	Unknown	High
P, M	*M*	*P*	–	*P*↔*M*	Low
P	*M*	*M*	+	*P* → *M*	High
P	*M*	*M*	–	*P* → *M*	Low
M	*P*	*P*	+	*P* ← *M*	High
M	*P*	*P*	–	*P* ← *M*	Low

Movable Genetic Elements Section

David J. Garfinkel

Thomas J. Burkett
Brenda Faiola
Sharon Moore
Lori Rinckel
Gil Sharon

The Movable Genetic Elements Section is committed to understanding the mechanism and consequences of Ty element retrotransposition in the baker's yeast *Saccharomyces cerevisiae*. We study the related elements Ty1 and Ty2, which are members of the widely dispersed Ty1-copia element superfamily. Throughout the report, the acronym Ty will refer to both Ty1 and Ty2 subfamilies. In the past year, we have continued our studies on Ty integration and the regulation of Ty-induced mutations by an unlinked suppressor gene called *SPT23*.

Homologous Recombination of Ty1 cDNA

The product of Ty reverse transcription is a linear cDNA intermediate that is associated with a Ty-virus-like particle (Ty-VLP). Ty1 cDNA can enter two distinct pathways that give rise to insertions into the genome: by the action of Ty1-integrase (IN) or by homologous recombination with endogenous Ty elements using cellular recombination machinery. The *his3*AI retrotransposition indicator gene, which we have described extensively in previous reports, has allowed us to study homologous recombination of Ty1 cDNA and pseudogene formation in yeast (also refer to the Genome Recombination, Rearrangement, and Regulation Section report). This indicator gene is useful because precise loss of the artificial intron (AI) and subsequent restoration of *HIS3* gene function can occur only if recombination of the marked Ty cDNA has proceeded through an RNA intermediate. In addition, since mitotic recombination of Ty1 cDNA depends on a variety of chromosomal genes, such as *RAD52*, and Ty1 transposition occurs in a *rad52* mutant background, we have distinguished Ty1 transposition from homologous recombination events by performing experiments in isogenic *RAD52* and *rad52* strains.

We have previously shown that Ty1m*his3*AI elements with a mutant IN domain carried on an inducible pGTy1 plasmid, which has been described extensively in previous reports, undergo high levels of cDNA recombination with both plasmid and chromosomal targets rather than IN-mediated transposition. These His^+ events are almost completely dependent on *RAD52* function, whereas wild-type Ty1 transposition events in this system are independent of *RAD52*.

The plasmid recombination events that result in a functional *HIS3* gene have been studied in greater detail by recovering them in *Escherichia coli* and determining their structure by restriction enzyme and polymerase chain reaction (PCR) analyses. In over half the cases, the only plasmid rearrangement observed was the conversion of the *his3*AI marker within the inducible Ty1 to a functional *HIS3* gene. These recombinants probably arise by homologous recombination between a marked Ty1*HIS3* cDNA and the plasmid Ty1m*his3*AI element. However, in about 20% of the plasmids that were checked (7/31), both *HIS3* and *his3*AI sequences were present on a single pGTy1 plasmid. Through further analysis of these recombinants, we found that six recombinants had two Ty1 elements inserted in a tandem head-to-tail configuration, with a single long terminal repeat (LTR) separating them: One element contained *his3*AI and the other contained *HIS3*. The final recombinant had three Ty1 elements in a similar head-to-tail configuration. These plasmid recombinants probably were generated by homologous recombination

between Ty1*HIS3* cDNA and the LTR of the resident Ty1m-*his3*AI element. We also observed multiple events in about 30% of the genomic recombinants (7/21). Therefore, our results indicate that when the IN function is blocked, complex multimeric Ty1 recombinants are much more common than was earlier thought. Perhaps higher levels of unprotected Ty1 cDNA accumulate in the IN mutant, allowing greater access to the homologous recombination machinery.

Are there evolutionary consequences of Ty1 cDNA recombination? It is striking that the optimum temperature for endogenous or *GAL1*-promoted Ty1 element transposition is 15°C, which is 15°C to 17°C below the optimum growth temperature for *S. cerevisiae*. Ty1-VLP reverse transcriptase (RT) activity is positively correlated with the temperature-sensitivity of Ty1 transposition. Ty1-VLPs prepared from cells grown at 20°C have higher levels of RT activity when assayed at low temperature than at high temperature, even though the levels of the p60 RT/ribonuclease H (RH) protein are normal in galactose-induced cells grown at high temperature.

To determine whether more Ty1 cDNA recombination events occur at the higher temperatures, we induced transposition in cells containing a wild-type pGTy1-H3m*his3*AI plasmid at 20°C or 32°C, then monitored Ty1*HIS3* cDNA recombination with the pGTy1-H3m*his3*AI plasmid. A dramatic temperature-dependent increase in the frequency of homologous recombination events mediated by Ty1*HIS3* cDNA was evident. These results indicate that Ty1 cDNA recombination plays a significant role in the life cycle and evolution of wild-type Ty elements. Homologous recombination of Ty cDNA provides another mechanism for disseminating Ty1 sequences while lowering the number of potentially deleterious de novo transposition events.

Biochemistry of Ty1 Integration; Purification and Initial Characterization of Recombinant Ty1-IN

Successful Ty1 retrotransposition requires Ty1-encoded functions necessary for protein maturation, Ty1-VLP assembly, reverse transcription, and transpositional integration. The Ty1-IN protein, which is required for de novo transposition, is expressed as part of a 190-kilodalton TYA1-TYB1 polyprotein. During Ty1-VLP maturation, the polyprotein is processed by Ty1- protease to form a mature IN polypeptide of about 90 kilodaltons. Investigation of the enzymology of the Ty1 integration reaction has been hampered by the lack of purified IN protein. Although Ty1-VLPs exhibit IN activity, they are not ideal sources of IN for two reasons: (1) The particles have been difficult to solubilize without complete loss of IN activity; and (2) determination of total Ty1-VLP protein concentration does not indicate the amount of any one protein present nor the degree of processing from immature to mature forms.

To study Ty1-IN in a form free from other Ty1-VLP proteins, we cloned the Ty1-H3 IN domain of *TYB1* into a galactose-inducible yeast expression vector. Upon induction, yeast cells containing the pGTy1-IN expression plasmid produced Ty1-IN of the expected molecular weight as compared to mature IN found in Ty1-VLPs (Figure 1). The expression plasmid was designed to allow protein purification by metal-chelate affinity chromatography. Using this procedure, we obtained fractions that are approximately

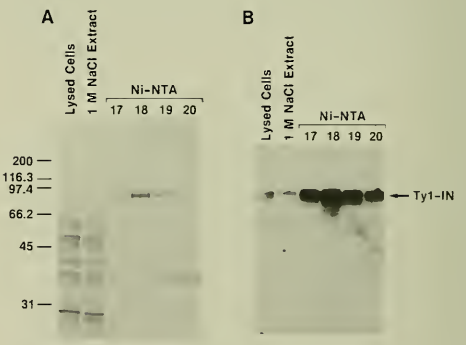


FIGURE 1. Purification of recombinant Ty1-IN protein from cells expressing pGTy1-IN. (A) Coomassie blue-stained SDS-polyacrylamide gel containing protein extracts from a yeast strain expressing pGTy1-IN. Numbers on the left represent molecular mass estimates in kilodaltons. Lane 1 contains a crude protein extract prepared by glass-bead lysis, lane 2 contains a 1 M NaCl extract of the lysed cellular pellet, lanes 3 to 6 contain fractions 17 to 20 eluted from an Ni-NTA metal-chelate column with increasing concentration of imidazole that have Ty1-IN activity and protein. (B) Immunoblot analysis of the gel shown in panel A using antibodies to Ty1-IN protein.

70% Ty1-IN protein, as judged by sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis and Coomassie blue staining. We confirmed that the major protein in these gels was Ty1-IN by immunoblot analysis using antiserum to Ty1-IN. Minor proteins in these fractions did not react with the IN antiserum, suggesting that these proteins were not aborted or degraded IN polypeptides.

We assayed Ty1-IN enzyme activity by monitoring the insertion of a double-stranded oligonucleotide substrate, derived from either the U3 or U5 LTR end of the Ty1 cDNA, into an identical oligonucleotide product. This assay has been used to detect IN activity from a variety of retroviruses, such as avian sarcoma virus, murine leukemia virus, and human immunodeficiency virus. Briefly, the assay requires IN to introduce a nick in the target molecule and transfer the 3′-OH end of the substrate to 5′-PO_4 of the nicked target. Radiolabeling the 5′ end of one of the strands allows the integration products to be visualized after electrophoresis on a denaturing sequencing gel and autoradiography. Integration results in a series of labeled molecules that are larger than the starting oligonucleotide. Formation of the longer molecules is accompanied by formation of shorter molecules that represent nicked targets from complete or incomplete integration events.

Using the oligonucleotide assay with a Ty1 U3 LTR oligonucleotide, we demonstrated a nicking and strand-transfer activity with recombinant Ty1-IN in various chromatographic fractions, which is similar to the activity observed with purified Ty1-VLPs (Figure 2). There appear to be certain sites in the target that are preferentially used for insertion. These preferential insertion sites result in more intensely labeled higher-molecular-weight products, while sites used less frequently result in fainter products. The combination of these sites generated a unique banding pattern of product that is conserved between recombinant and Ty1-VLP–associated IN. This result suggests that recombinant Ty1-IN mimics Ty1-VLP–associated IN protein.

We then compared the activity of recombinant Ty1-IN on U3 and U5 LTR oligonucleotides that were labeled to the same specific activity with [γ-^{32}P]dATP and T_4 polynucleotide kinase (Figure 3). In this experiment, both recombinant Ty1-IN and Ty1-VLPs showed identical product patterns with wild-type U3 and U5 LTR oligonucleotides. As expected, there were characteristic differences in the

FIGURE 2. DNA strand transfer activity of Ty1-IN protein. NP, no protein added to the reaction; VLP, IN activity present in Ty1-VLPs; Ni-NTA 17 to 20, IN protein fractions 17 to 20 from the column described in Figure 1.

FIGURE 3. Relative activities of Ty1-IN protein for the U3 and U5 LTR ends. NP, no protein; IN, recombinant Ty1-IN protein; VLP, Ty1-VLPs.

integration pattern using U3 versus U5 LTRs. In contrast to retroviral IN-catalyzed reactions, our results show that Ty1-IN protein does not process any nucleotides from the substrate prior to integration. These intermediates would appear as prominent bands that are slightly shorter than the full-length oligonucleotide. This result provides direct support for the previously held idea that a Ty1 element integrates as a blunt-ended molecule. Finally, it has been reported by others that retroviral IN proteins are more active for end-processing and strand transfer when the U5 LTR oligonucleotide substrate end is used than when the U3 LTR end is used in the reaction. Our initial results suggest that Ty1-IN does not show a preference for one end over the other for strand transfer.

We are interested in chemical or protein factors that stimulate Ty integration in vitro or in vivo. Therefore, we used the oligonucleotide assay to determine whether the metal cofactors Mg^{2+} or Mn^{2+} could stimulate IN activity (Figure 4). The results show that although Ty1-IN protein is active in the presence of Mg^{2+}, the amount of product formed with Mn^{2+} is about threefold greater than with Mg^{2+}. The number of products smaller than the substrate was also greatly decreased in the presence of Mg^{2+} compared to that observed with Mn^{2+}. This result suggests that nicking of the target may be the rate-limiting step in the reaction.

In the coming year, further biochemical and genetic characterization of Ty1-IN will be a major focus of the Movable Genetic Elements Section. We plan to develop other assays to measure integration activity, such as the ability of Ty1-IN to catalyze insertion of a full-length Ty1-LTR into a heterologous target or the ability of Ty1-IN to catalyze a reverse reaction called disintegration. We will determine the minimal sequence required for integration, using mutant oligonucleotides in the assay described above. Although the activity that we have purified is clearly enriched for Ty1-IN protein, our preparations are only about 70% pure. Therefore, we will purify Ty1-IN to homogeneity by additional chromatographic procedures. Further purification of recombinant Ty1-IN will be essential for structural studies as well as for determining whether cellular cofactors are required for or stimulate IN activity in vitro. Since we have chosen to express Ty1-IN in yeast, these endogenous factors, if present, may copurify with IN.

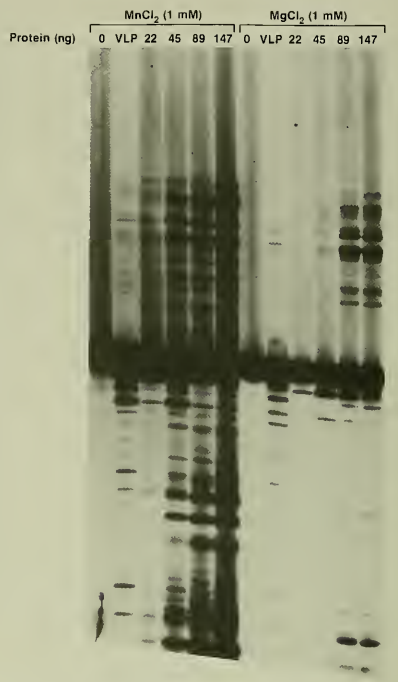


FIGURE 4. Metal cofactor requirements of Ty1-IN protein. Integration assays were performed in the presence of either 1 mM $MnCl_2$ or 1 mM $MgCl_2$. Lanes marked 0 and VLP contain no protein and Ty1-VLPs, respectively. The remaining lanes contain serial dilutions of recombinant Ty1-IN protein.

Integration Specificity of Ty Elements

Previous studies of de novo Ty transposition events, as well as retroviral insertions, have shown that these elements do not insert randomly in the genome. Ty elements show a strong insertion-site preference for genomic regions 5′ to open reading frames, within 1 kilobase of tRNA genes, other Ty elements, or solo LTRs. A variety of studies have led to the view that higher-order structure of the genome dictates retrotransposon and retrovirus insertion specificity. None of these studies, however, has biochemically defined the components of chromatin that are involved in Ty targeting.

We are studying Ty insertion-site specificity at the yeast *CAN1* locus in strains that have altered chromatin structure. Null mutations in *CAN1*, which encodes a high-affinity arginine permease, confer resistance to the arginine analog canavanine. In the inducible pGTy1-H3m*his3*AI element transposition system, over 70% of the Can^r, His$^+$ cells contain Ty-induced mutations at the *CAN1* locus. Using PCR amplification, we analyzed total DNA from these mutants to determine whether a Ty element has been inserted at *CAN1*, which type of element is present, and what are its orientation and map position. Last year, we presented the pattern of Ty2 element insertions at *CAN1* in a wild-type strain with a mapping resolution of 50 base pairs.

Over the past year, we compared the pattern of wild-type Ty insertion to that obtained in isogenic strains that have lower levels of the histone proteins H2A and H2B. This comparison was achieved by using strains containing a deletion of either the *TRT1* or *TRT2* locus. Each of these loci encodes the H2A and H2B proteins; however, the loci are regulated differently. *TRT1* is regulated by the level of H2A and H2B in the cell, while *TRT2* is constitutively expressed and is not regulated by the level of H2A and H2B. Strains lacking *TRT1* have a variety of cellular defects and have more nucleosome-free regions in the genome; strains lacking *TRT2* have only minor changes in cellular processes and have normal nucleosome phasing.

We predicted that Ty insertion-site preference would be altered in a *trt1*Δ mutant background. After analyzing over 700 independent Ty transposition events at *CAN1*, we observed a striking change in the insertion pattern when *TRT1* was deleted (Figures 5 and 6). Chi-square analysis of these data shows a significantly different insertion pattern ($P \leq 0.0001$) in the 0 to +500 and +1550 to +1950 nucleotide intervals when the *trt1*Δ strain is compared to the wild type or the *trt2*Δ strain ($P = 0.35$). To control for selection bias in the experiment, we introduced the *trt1*Δ deletion into several Ty-induced Can^r mutants isolated in a *TRT1* strain and we introduced the wild-type *TRT1* gene into several Can^r mutants isolated in a *trt1*Δ mutant background; we then tested these strains for resistance to canavanine. In particular, Ty insertions that mapped in the affected regions were chosen for this analysis. The Can^r phenotype remained the same in all of these strains, indicating that selection bias cannot account for the alterations observed in a *trt1*Δ mutant. Taken together, these results suggest

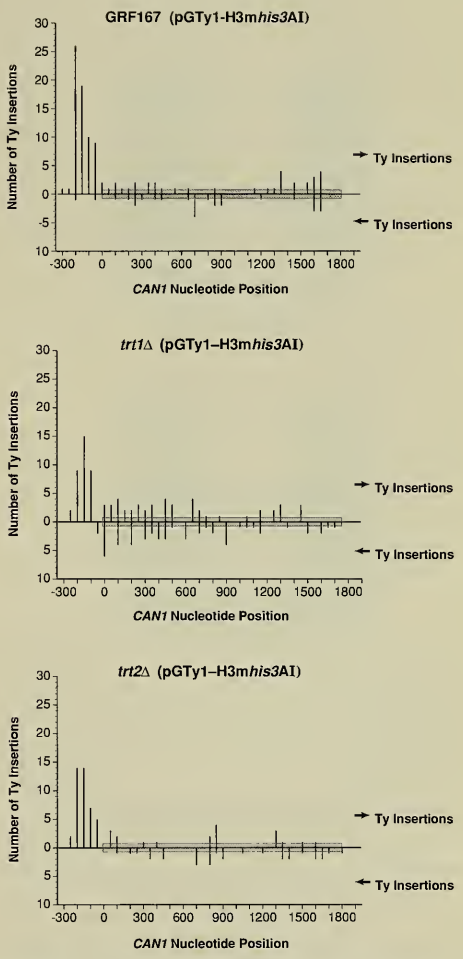


FIGURE 5. Histogram of independent Ty insertions at the *CAN1* locus in strains GRF167 (*TRT1*, *TRT2*), *trt1*Δ (*trt1*Δ, *TRT2*), and *trt2*Δ (*TRT1*, *trt2*Δ) following galactose induction of pGTy1-H3m*his3*AI transcription. The boxed segment indicates the *CAN1* coding region. Ty elements that have inserted in the same transcriptional orientation as *CAN1* are indicated by →; elements in the opposite orientation are indicated by ←. The insertions were mapped by PCR analysis of genomic DNA prepared from independent Can^r isolates, using Ty1-, Ty2-, and *CAN1*-specific oligonucleotides.

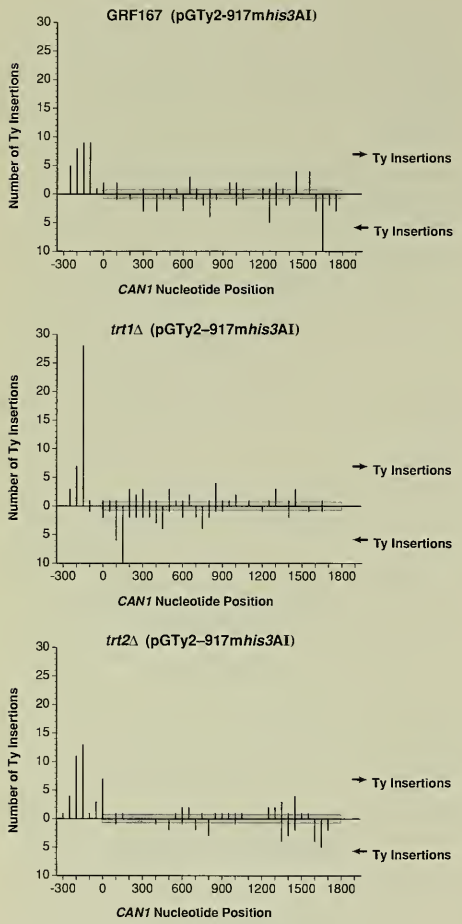


FIGURE 6. Histogram of independent Ty insertions at the *CAN1* locus in strains GRF167 (*TRT1, TRT2*), *trt1*Δ (*trt1*Δ, *TRT2*), and *trt2*Δ (*TRT1*, *trt2*Δ) following galactose induction of plasmid pGTy2-917m*his3*AI transcription. Refer to Figure 5 for more information.

that nucleosome density is an important factor in determining Ty1 and Ty2 element target-site preference. In the coming year, we will determine the nucleosome phasing in these regions of *CAN1* in the wild type and the *trt* mutants to determine whether a change in nucleosome phasing is correlated with the altered insertion pattern observed in these intervals.

In the course of these studies, we sequenced a total of 14 Ty1/*CAN1* and 31 Ty2/*CAN1* insertion-site junctions in wild-type cells. Based on the 5-base pair duplication generated by Ty transposition, the Ty1 insertions gave a consensus sequence of G/A/C-T/A-T/A-T/A-T/G and the Ty2 insertions gave G/A/C-T/A-T/A-T-C/G/T. These consensus sequences are similar to the consensus sequence obtained from previous studies using predominantly Ty1 insertions (G/A/C-X-A/T-X-T/G/C). Thirty-eight Ty1 and Ty2 insertions in the +1550 to +1950 interval were sequenced from the wild-type strain. Interestingly, 9 of these mapped precisely to position +1599 and 12 more mapped to position +1617 in the *CAN1* gene. In addition, no orientation bias was observed. These results suggest that Ty target-site preference can be quite precise, although this specificity does not appear to be determined simply by the sequence of the target.

Regulation of Ty-Induced Mutations by *SPT23*

Mutations caused by insertion of a Ty element in the promoter region of a target gene are generally unstable. Reversion of these mutations is caused either by utilization of cis-acting sequences located in the adjacent Ty element or by mutation of unlinked trans-acting suppressor genes. In previous years, we summarized our work on a suppressor gene called *SPT23*. Over the past year, we have studied the possible interactions between *SPT23* and genes involved in the maintenance and alteration of chromatin structure.

We isolated *SPT23* as a multicopy suppressor of the Ty-induced mutations *his4-912*δ and *lys2-61*. Additional analyses demonstrated that *SPT23* suppresses a variety of Ty-induced mutations. We have presented the molecular description of *SPT23* in previous reports. Briefly, *SPT23* is predicted to encode a slightly basic, asparagine- and serine-rich, 68-kilodalton protein with no known homologs. Disruption mutants of *SPT23* are viable, indicating that it is not essential for growth under standard laboratory conditions. Interestingly, correct expression of the conditional Ty-induced mutations *his4-912*δ and *lys2-61* at the permissive temperature requires *SPT23*.

Recently, Fred Winston's laboratory (Harvard Medical School) has shown that the sucrose nonfermenting (*SNF*) genes *SNF2*, *SNF5*, and *SNF6*, which regulate expression of the invertase gene *SUC2*, also affect *his4-912*δ expression. Like *spt23* mutants, strains containing *his4-912*δ and either *snf2*, *snf5*, or *snf6* are unconditional histidine auxotrophs. There are also genes that, when mutated, restore proper *his4-912*δ expression in a wild-type and an *snf2*, *snf5*, or *snf6* mutant background. One of these genes is *SPT6*(*SSN20*), which was isolated in His^+ or Suc^+ reversion analyses. As in *snf2 snf5* double mutants, histidine prototrophy was restored in *spt23 spt6* double mutants. Although there are definite similarities between *SPT23* and the *SNF* genes, specific phenotypic tests have suggested that *SPT23* is not an *SNF* gene.

Even though *SPT23* is probably not an *SNF* gene, we observed that multicopy expression of *SPT23* suppressed the defect imposed on *his4-912*δ expression by an *snf2* mutant (Figure 7). Interestingly, it was recently reported that mutations in *SNF2* prevent removal of chromatin components from the *SUC2* promoter region that represses *SUC2* transcription (Hirschorn et al., Genes Dev 6:2288, 1993). Combined with our data, this result suggests that a similar mechanism is responsible for the inability of *his4-912*δ to be properly expressed in an *snf2* mutant background. Furthermore, multicopy expression of *SPT23* may remove the chromatin components that repress normal transcription of *his4-912*δ. In the coming year, we will test this model by determining whether *SPT23* multicopy expression alters the chromatin structure of the *his4-912*δ promoter.

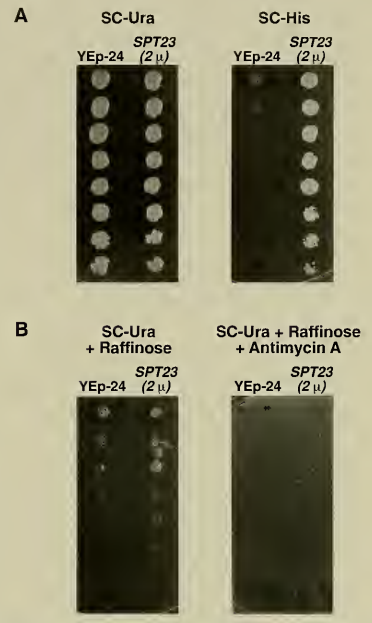


FIGURE 7. *SPT23*-mediated multicopy suppression of a *his4-912*δ, *snf2* mutant. (A) Yeast strains containing the *his4-912*δ and *snf2* mutations and either *SPT23* on a multicopy plasmid (2 μ) or a vector control (YEp-24) were grown in synthetic complete medium lacking uracil (SC-Ura), serially diluted, and spotted onto SC-Ura or synthetic complete medium lacking histidine (SC-His). Plates were incubated at 30°C for 5 days. (B) Lack of *SPT23*-mediated multicopy suppression of *SUC2* in an *snf2* mutant. The same strains as in panel A were spotted onto SC-Ura medium with raffinose as the carbon source in the absence (left) or presence (right) of the respiratory inhibitor antimycin A. In a parallel experiment, we showed that an isogenic *SUC, SNF* strain grew on SC-Ura medium containing antimycin (data not shown).

Chemistry of Carcinogenesis Laboratory

Anthony Dipple

Robert C. Moschel

The Chemistry of Carcinogenesis Laboratory investigates the mechanisms through which chemical carcinogens initiate the carcinogenic process. Largely, research effort is focused on the chemical modification of DNA by reactive carcinogenic chemicals, the repair of such damage, and the conversion of unrepaired damage to mutagenic events. The agents that are studied most extensively are carcinogenic alkylating agents and polycyclic aromatic hydrocarbons. DNA adducts generated by these agents are explored through both chemically and biologically oriented approaches to determine how specific carcinogen-DNA adducts might give rise to a particular biological effect.

The Molecular Aspects of Chemical Carcinogenesis Section, headed by Anthony Dipple, is particularly concerned with the hydrocarbon carcinogens, the chemical mechanisms that determine their modification of DNA, and the mutational spectra generated by related hydrocarbon dihydrodiol epoxide metabolites. In this report, we summarize studies on the chemistry of adenine aralkylation, the sequence specificity of dihydrodiol epoxide reactions with DNA, the mutagenic specificities of nitric oxide and some hydrocarbon derivatives, and the in vitro action of several DNA polymerases on oligodeoxyribonucleotides containing single benzo[*a*]pyrene-deoxyadenosine adducts at known sequence locations.

The Carcinogen-Modified Nucleic Acid Chemistry Section, headed by Robert C. Moschel, combines synthetic organic chemistry and molecular biological techniques to prepare oligodeoxyribonucleotide substrates containing specific carcinogen-modified bases and to evaluate the effects of DNA repair on these substrates and on the mutagenic potency of the carcinogen-modified bases. In collaborative studies, this section has discovered agents that inactivate O^6-alkylguanine-DNA alkyltransferase (AGT) and is seeking to develop these agents as chemotherapeutic adjuvants that can substantially enhance the anticancer action of chloroethylating drugs. In this report, the modulating effect of AGT on the biological potential of O^6-substituted guanines in codon 12 of H-*ras* sequences is described, as are the evaluation and synthesis of 13 new AGT inhibitors. The effectiveness of O^6-benzylguanine used in combination with 1,3-*bis*(2-chloroethyl)-1-nitrosourea for human brain tumors, treated as xenografts in nude mice, is also described.

Molecular Aspects of Chemical Carcinogenesis Section

Anthony Dipple

Rajiv Agarwal
Karen A. Canella
Donna F. Christner
John E. Page
Chengyi Qian
Helen L. Ross
Michael N. Routledge
Jan Szeliga

The research of the Molecular Aspects of Chemical Carcinogenesis Section is directed at understanding the mechanisms through which organic chemical carcinogens initiate the carcinogenic process. A key step in this process is the covalent modification of cellular DNA by a reactive carcinogen metabolite. Therefore, we are interested in both the chemistry of such carcinogen-DNA interactions and the mechanisms through which carcinogen-DNA adducts lead to biological responses, such as mutagenesis and carcinogenesis. In all of our studies, we have focused on the polycyclic aromatic hydrocarbon carcinogens, although we have also used various model compounds. In this report, we describe ongoing studies dealing with the mechanism of DNA aralkylation, the sequence specificity of hydrocarbon dihydrodiol epoxide-DNA reactions, the mutagenic specificity of nitric oxide and reactive hydrocarbon derivatives, and polymerase action on specific hydrocarbon carcinogen-nucleic acid adducts.

Our earlier chemical studies generated insight into the factors that determine site selectivity in guanine alkylation and aralkylation in DNA. Currently, we are investigating the factors that determine reaction sites on adenine residues in DNA. These studies are revealing some differences from, as well as similarities to, the guanine aralkylation mechanism. We have also extended our studies of the sites of reaction of hydrocarbon carcinogen metabolites within DNA sequences. These studies indicate that although closely related hydrocarbon dihydrodiol epoxides exhibit similarities in their reactions with DNA, their overall distributions through a specific sequence are different. Mutagenic specificity studies for reactive hydrocarbon derivatives have been extended to include 7-bromomethylbenz[*a*]anthracene and 7-bromomethyl-12-methylbenz[*a*]anthracene. In the pS189/*supF* system, 7-bromomethylbenz[*a*]anthracene-DNA adducts were more mutagenic than adducts from the 12-methyl compound. In collaboration with Larry K. Keefer and David A. Wink (Laboratory of Comparative Carcinogenesis, NCI-FCRDC), we have investigated the nature of nitric oxide-induced mutations generated within human cells. In contrast to our findings with hydrocarbons, nitric oxide predominantly induced transition mutations, and AT→GC changes outnumbered GC→AT changes. Through ongoing collaborations with Donald M. Jerina (National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD), we have obtained a series of oligodeoxyribonucleotides containing a single benzo[*a*]pyrene-deoxyadenosine residue in a specific location. These substrates have allowed us to substantially extend the scope of our studies of polymerase action on individual hydrocarbon-deoxyribonucleoside adducts in oligodeoxyribonucleotides.

Chemistry of Adduct Formation

To investigate the mechanism of aralkylation at the 1- and N^6-positions of adenine residues, we determined the yield and stereochemistry of products generated from reaction with optically active styrene oxide.

The reaction of racemic styrene oxide with adenosine potentially yields eight styrene oxide-adenosine adducts. These adducts arise from the reaction of the two optical isomers of the epoxide at either the α- or β-carbon with either the 1- or N^6-position of the adenine residues.

Figure 1 illustrates the eight putative adducts, along with their stereochemical interrelationships and the routes from which these adducts were obtained. The N^6-substituted adenosines were produced from reactions of the appropriate amino compounds with 6-chloropurine riboside. In the case of N^6-products formed at the α-carbon of the epoxide, we used commercially available, optically active 2-amino-2-phenylethanols in separate reactions to prepare the diastereomeric adducts (labeled N^6-αI and N^6-αII in Figure 1). Since displacement of the chlorine by the purine amino group did not involve breaking any bonds at the chiral α-carbon, the products were assigned the same absolute configuration as the starting 2-amino-2-phenylethanol.

Although we prepared the 1-substituted adenosines from racemic styrene oxide, we also carried out reactions with the individual enantiomers (Figure 1). Since formation of the product at the epoxide's β-carbon did not change the configuration at the chiral α-carbon, the two products with a linkage between the β-carbon and the 1-position of adenosine (labeled 1-βI and 1-βII) were assigned the configuration of the styrene oxide enantiomer from which they were formed. The configuration of the 1-substituted

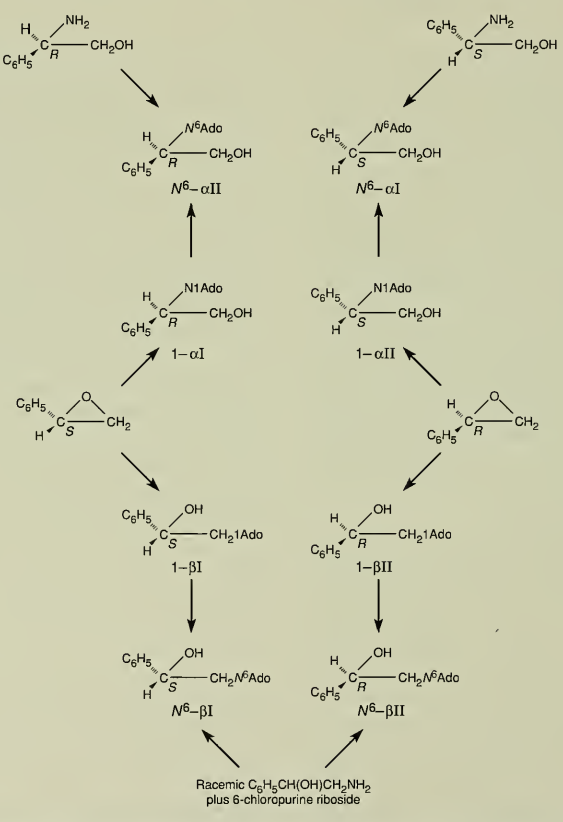


FIGURE 1. Chemical relationships and interconversions through which the absolute stereochemistries of styrene oxide-adenosine products were defined.

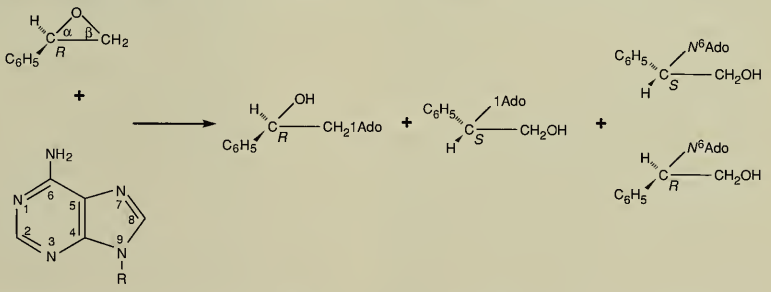
products formed at the α-carbon of the epoxide (labeled 1-αI and 1-αII) was established by subjecting these products to alkali-catalyzed Dimroth rearrangement to the corresponding N^6-substituted products, which had known configurations. Circular dichroism spectra and high-pressure liquid chromatography retention times were used to establish identity. Thus, we found that the major 1-substituted adenosine produced from reaction with the α-carbon of an optically active styrene oxide had undergone inversion of configuration at the chiral center.

Similarly, we established the configurations of the N^6-substituted adenosines with linkage to the β-carbon of styrene oxide (labeled N^6-βI and N^6-βII in Figure 1) by separately subjecting 1-βI and 1-βII, which had known configurations, to Dimroth rearrangement. The Dimroth rearrangements of the four 1-substituted adenosines also served to confirm their assignments as products of reaction at either the α- or β-carbon of the styrene oxide, since the structures of the N^6-α and N^6-β products were established by their synthesis from 6-chloropurine riboside. The four pairs of diastereomeric products were labeled I or II based on their sequence of elution from the chromatographic system used for their separation.

To quantitate and evaluate the stereochemistries of products of styrene oxide-adenosine reactions, we carried out reactions in neutral aqueous solutions containing [^{14}C]adenosine and optically active styrene oxide for 24 hours. Thereafter, we added quantities of adenosine and the eight potential products (Figure 1) in amounts that were sufficient for their detection by ultraviolet light absorbance during subsequent chromatography. By determining the radioactivity in each peak, we were able to quantitate the various products. A summary of the stereochemical results obtained from duplicate analyses for each styrene oxide enantiomer is presented in Table 1.

Reactions forming 1-substituted adenosines were stereospecific, since only the retained configuration was found in the products at the β-carbon and only the inverted configuration was found in the products at the α-carbon. These results are consistent with those expected for a reaction occurring through an S_N2 mechanism. Also consistent with an S_N2 mechanism is the presence of both α- and β-substituted products, with the β-substituted products being formed in greater yield. Simple S_N2 displacements on epoxides usually occur preferentially at the less-branched carbon, i.e., the β-carbon atom in styrene oxide.

TABLE 1. Products from optically active styrene oxides and adenosine*



	S/R	*S/R*	*S/R*
(*R*)– Styrene Oxide	0.0/0.31	0.06/0.0	0.42/0.20
	0.0/0.31	0.08/0.0	0.47/0.12
(*S*)– Styrene Oxide	0.31/0.0	0.0/0.11	0.16/0.39
	0.30/0.0	0.0/0.13	0.12/0.41

*Numbers represent the fraction of total adenosine products formed. Conversion of adenosine to product averaged 2.7% and 2.6% for the two reactions of the (*R*) and (*S*) enantiomers, respectively.

The situation for the N^6-substituted adenosines is quite different. Clearly, in the case of the α-substituted styrene derivatives, which represent the major adenosine aralkylation products, derivatives with either retained or inverted stereochemistry were formed. Although inversion of configuration predominated over retention, the difference between the findings for the N^6- and 1-substituted products suggests that a more ionic character is involved in the reactions leading to aralkylation at the N^6-site. We are therefore surprised that the yield of the N^6-substituted product is greater than that of the 1-substituted product. However, further investigations with this model system may help us to understand why some biologically potent hydrocarbon dihydrodiol epoxides react so extensively with the N^6-site on adenine residues in DNA.

Sequence Specificity Studies

Dihydrodiol epoxides of polycyclic aromatic hydrocarbons react with DNA at purines to form bulky adducts that act as blocks to transcription and replication. DNA polymerase arrest can be used, therefore, to determine the sites of adduct formation within a particular sequence. We previously reported that the four configurational isomers of benzo[*c*]phenanthrene dihydrodiol epoxide have unique polymerase arrest spectra in the *supF* gene of the shuttle vector pS189 (1992 Annual Report). We have recently extended these studies to determine the sites of adduct formation in the *supF* gene by (+) *anti* benzo[*a*]pyrene 7,8-dihydrodiol 9,10-epoxide, (±) *anti* 5-methylchrysene 1,2-dihydrodiol 3,4-epoxide, and (±) *anti* 7-methylbenz[*a*]anthracene 3,4-dihydrodiol 1,2-epoxide.

Figure 2 shows the results of a polymerase arrest assay using pS189 DNA treated with increasing concentrations of (±) *anti* 5-methylchrysene dihydrodiol epoxide and incubated with Sequenase (version 2.0), end-labeled primer, and deoxyribonucleoside triphosphates. Whereas the control lanes (i.e., unreacted DNA) contained mostly high-molecular-weight material at the top of the gel, which is indicative of unhindered extension of the primer, the lanes with treated DNA contained many bands resulting from polymerase arrest. Variation of carcinogen dose did not change the pattern of polymerase arrest; however, with decreasing dose, the intensity of all bands generally decreased. With increasing carcinogen dose, the probability of the polymerase encountering an adduct increased and, thus,

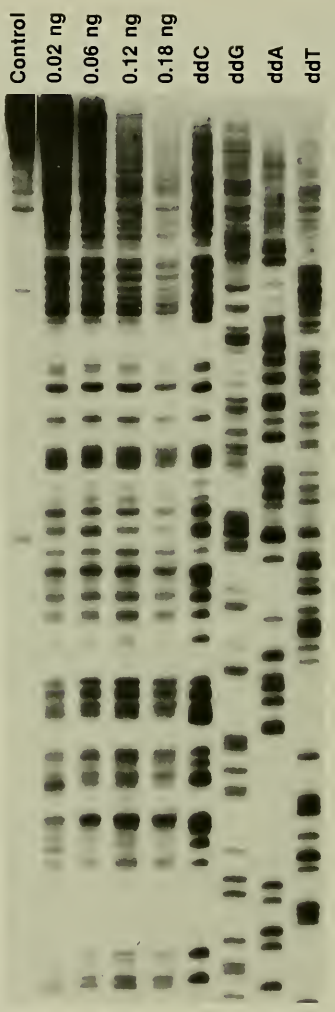


FIGURE 2. Autoradiograph of a polymerase arrest assay using Sequenase version 2.0 and the *supF* gene of pS189 treated with *anti* 5-methylchrysene 1,2-dihydrodiol 3,4-epoxide. DNA (10 µg) was chemically modified with 0.02, 0.06, 0.12, or 0.18 ng of the dihydrodiol epoxide. The control lane shown on the left contains DNA treated with buffer only. The four lanes on the right contain the dideoxynucleotide sequencing reactions.

the amount of high-molecular-weight material at the top of the gel decreased. The arrangement of the arrest bands reflected the pattern of the bands in the adjacent dideoxycytidine sequencing lane, indicating that most of the polymerase arrest was at the guanine residues in the template. By densitometry, we estimated the intensity of the polymerase arrest bands in a given lane and then compared such traces with sites of mutation hotspots determined either in our laboratory (5-methylchrysene and 7-methylbenz[*a*]anthracene dihydrodiol epoxides) or in Veronica Maher's laboratory (benzo[*a*]pyrene dihydrodiol epoxide; Yang et al., Proc Natl Acad Sci USA 84:3787, 1987) (Figure 3).

In Figure 3, the vertical bars represent the position and intensity of the polymerase arrest bands (as a percentage of all bands in that sequence) and the boxes between the strands represent the sites of mutation hotspots. Although the dihydrodiol epoxides that we examined were similar in that they all preferentially reacted with guanine residues in DNA, their polymerase arrest spectra were not identical. Further, within the limits of the *supF* sequence, it was possible to examine the sequence contexts that led to the most intense polymerase arrest. For benzo[*a*]pyrene dihydrodiol epoxide, the most intense arrest bands appeared at the underlined G in the sequence GGG, whereas for the two other compounds the preferred sequence context was (C or T)GG. This result suggests that hydrocarbon structure plays a role in determining the sequence specificity of DNA modifications.

Mutation Studies

Nitric Oxide. Nitric oxide (NO) is an important molecule that is mutagenic to *Salmonella typhimurium* and human cells in culture and has been proposed as a candidate endogenous mutagen because of its ability to deaminate DNA. We have used the *supF* forward mutation assay to study mutations induced in the *supF* gene following treatment of the shuttle vector, on which the *supF* gene is located, with an aqueous solution of NO gas, a spermine-NO complex (sperNO) that breaks down in solution to release NO in a controlled manner, or nitrite. As with ongoing polycyclic aromatic hydrocarbon mutagenicity experiments, we have used the recently modified shuttle vector of Parris and Seidman, pSP189. This plasmid contains a randomly generated signature sequence 3′ to the *supF* gene, which allows us to determine whether plasmids with identical mutations have arisen as a result of independent mutations or are siblings of a single mutated plasmid. Using pSP189 for the NO and recent polycyclic aromatic hydrocarbon studies, we found that, as shown by Parris and Seidman for ultraviolet light-induced mutations, siblings constitute less than 1% of sequenced mutants.

In the case of the aqueous solution of NO gas, the most common mutations were AT→GC transitions, which would be consistent with deamination of deoxyadenosine (Table 2). The types of mutations and, to some extent, their distribution, were similar regardless of whether the treated plasmid was replicated in human Ad293 cells or in *Escherichia coli* MBM7070 cells (Figure 4). This similarity suggests that cellular systems of repair and replication in these two very different cell types did not define the types of mutations generated.

For the sperNO compound, mutations occurred primarily at GC pairs, with GC→AT mutations being the most common (Table 2). This surprising contrast between results with NO gas and sperNO suggests either that the delivery system of NO is important in defining mutations or that the spermine component of the compound influences the reaction of NO with the plasmid.

Nitrite is the autoxidation product of NO in solution. Mutation frequency for nitrite at pH 7.4 was about tenfold

TABLE 2. Types of induced mutations

	Percentage of mutations			
Base change	NO (Ad293)	NO (*E. coli*)	SperNO	Nitrite (pH 5.4)
Transversion				
GC→TA	4	0	16	29
GC→CG	1	1	11	9
AT→CG	1	1	3	0
AT→TA	1	0	4	0
Transition				
GC→AT	28	22	56	43
AT→GC	61	71	5	12
Base deletion	1	0	1	3
Large deletion	1	5	3	2
Total mutations sequenced	71	78	73	98

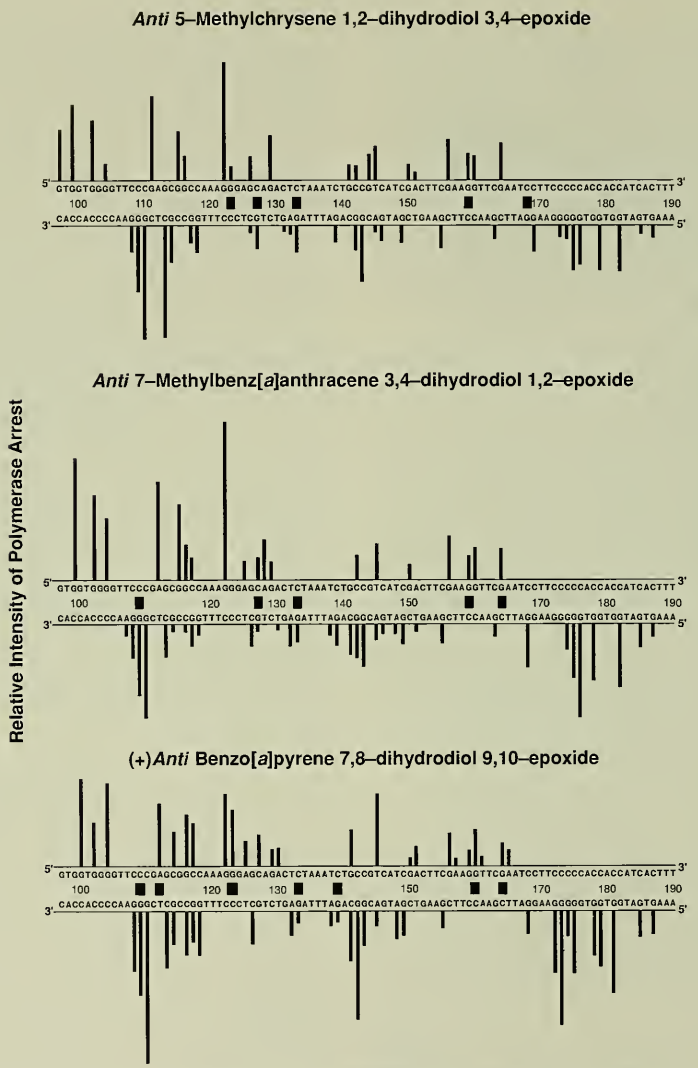


FIGURE 3. Sites of polymerase arrest in the *supF* gene from DNA treated with *anti* 5-methylchrysene 1,2-dihydrodiol 3,4-epoxide, *anti* 7-methylbenz[*a*]anthracene 3,4-dihydrodiol 1,2-epoxide, or *anti* benzo[*a*]pyrene 7,8-dihydrodiol 9,10-epoxide. Sequence numbering is indicated between the two strands. Vertical bars represent the sites and relative intensities of polymerase arrest, and boxes between the two strands represent the sites of mutation hotspots. Mutation data were derived from our previously published results except for *anti* benzo[*a*]pyrene 7,8-dihydrodiol 9,10-epoxide (Yang et al., Proc Natl Acad Sci USA 84:3787, 1987).

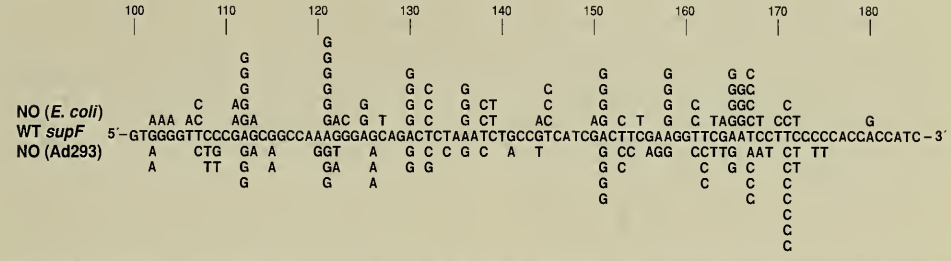


FIGURE 4. Distribution of NO-induced point mutations in the *supF* gene. Shown above the wild-type *supF* sequence are mutations arising when the shuttle vector was replicated solely in *E. coli* MBM7070 cells. Shown below are mutations arising after replication in Ad293 cells. Only mutations within the structural *supF* gene are shown.

lower than that for NO, but mutation frequency for nitrite increased dramatically with decreasing pH. We sequenced the mutants generated after treatment of the plasmid with 1 M nitrite at pH 5.4 and found that mutations occurred mainly at GC pairs (Table 2).

For each of the three treatments discussed above (i.e., sperNO, nitrite, and an aqueous solution of NO gas), a unique mutational spectrum was obtained. In all cases, the dominant mutations induced in these experiments can be explained by a deamination mechanism. We hope that chemical analysis of the products of these treatments will help to explain the different mutation specificities that we have found.

Hydrocarbon Derivatives. To extend our investigations of the dependence of mutagenic specificity on chemical structure, we examined the mutational spectra for two synthetic polycyclic aromatic hydrocarbon derivatives that are closely related, 7-bromomethylbenz[*a*]anthracene and 7-bromomethyl-12-methylbenz[*a*]anthracene. These compounds are direct-acting carcinogens that react with DNA through the bromomethyl group to yield products on the exocyclic amino groups of the bases in the following order: guanine > adenine > cytosine. Table 3 lists induced *supF* point mutation frequencies for the bromomethyl derivatives, showing that 7-bromomethylbenz[*a*]anthracene is the more mutagenic of the two compounds. The 7-bromomethyl derivative reacts more extensively with DNA than does the 12-methyl derivative (three- to fivefold). However, the tenfold difference in mutation frequency suggests that, in the pS189 system, a 7-bromomethylbenz[*a*]anthracene lesion is intrinsically more mutagenic than a 7-bromomethyl-12-methylbenz[*a*]anthracene lesion.

The types of point mutations induced by these hydrocarbon derivatives are illustrated in Table 4. Approximately the

TABLE 3. Induced *supF* point mutations

Treatment	Dose (nmol/ml)	Point mutants	Frequency of point mutations x 10^4	Mutation frequency nmol/ml
Solvent only	–	3	0.09	–
7-BrMeBA	8.40	108	24.40	2.90
7-BrMe12MeBA	13.40	103	3.90	0.29

TABLE 4. Point mutations induced by bromomethylbenz[*a*]-anthracenes

Type of mutation	Percentage of mutations 7-BrMeBA	7-BrMe12MeBA
Mutations at GC pairs		
GC, TA	56	60
GC, CG	3	15
GC, AT	7	8
Mutations at AT pairs		
AT, TA	3	3
AT, CG	2	1
AT, GC	5	1
(+1) Insertion	3	6
(–1) Deletion	20	7
Total mutations sequenced	123	129

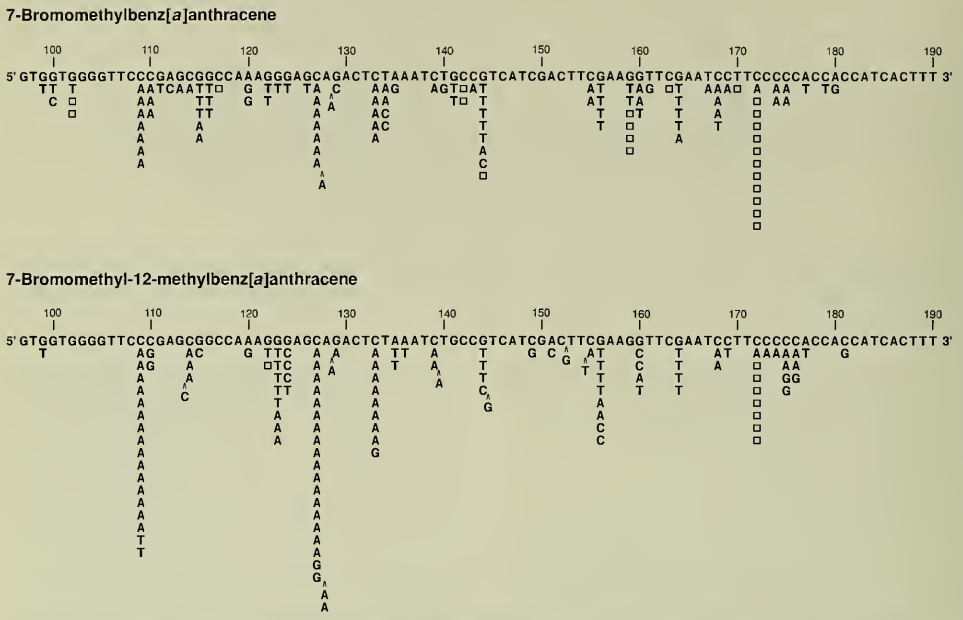


FIGURE 5. Distribution of 7-bromomethylbenz[*a*]anthracene– and 7-bromomethyl-12-methylbenz[*a*]anthracene–induced point mutations in the *supF* tRNA gene. Mutations are listed below the wild-type *supF* sequence for each carcinogen; ^, insertion; □, deletion. Deletion mutations occurring in a sequence of two or more identical bases are arbitrarily assigned to the 5′-base.

same number of mutations was analyzed for each compound. The most common base substitutions were at GC pairs and involved a G→T transversion. Aside from G→T transversions, a G→C transversion was the most prevalent mutation for the 12-methyl derivative and deletion of a GC pair was the most prevalent mutation for the 7-bromomethyl derivative. All of the single base deletions involved a GC pair and usually occurred at sites where two or more guanines were adjacent to one another.

The mutational spectra for both hydrocarbons are shown in Figure 5. All mutation hotspots occurred at GC pairs, and three hotspots were common to both compounds (sites 109, 127, and 133). 7-Bromomethylbenz[*a*]anthracene mutations were more scattered along the *supF* gene than were those generated by the 12-methyl derivative, and the two spectra were different according to the statistical test of Adams and Skopek ($p = 0.02$) (J Mol Biol 194:391, 1987).

These studies showed that a simple chemical modification (i.e., a methyl group in the bay region) can change mutagenic specificity and lead to a reduction in mutagenic activity without substantially changing the type of mutation preferentially induced (in this case, G→T transversion).

Polymerase Studies

Our most recent study of DNA polymerase action on specific hydrocarbon adducts at defined sites in oligodeoxyribonucleotides is focused on two oligodeoxyribonucleotide sequences from the *supF* gene with benzo[*a*]pyrene adenine adducts placed at the fourth nucleotide from the 5′-end. In the present study, we examined the four trans deoxyadenosine adducts derived from the four configurationally isomeric benzo[*a*]pyrene dihydrodiol epoxides. The sequences 5′-TTTAGAGTCTGCTCCC-3′ and

5′-CAGATTTAGAGTCTGC-3′ are from the *supF* gene, which we used as a target in our mutation studies, and represent adduct sites of high and low mutagenic activity, respectively, when benzo[*c*]phenanthrene dihydrodiol epoxide is used as a mutagen. The sequences are referred to herein as the hotspot and coldspot templates, respectively.

We have examined three different polymerases: Sequenase version 2.0, the Klenow fragment of DNA polymerase I, and the reverse transcriptase of human immunodeficiency virus type 1 (HIV-1), provided by Stephen Hughes (Gene Expression in Eukaryotes Section, Molecular Mechanisms of Carcinogenesis Laboratory). With Sequenase, all four adducts showed incorporation of the correct base 3′ to the adduct followed by polymerase arrest. At longer incubation times and with higher nucleotide concentrations, there was some incorporation of adenine across from the adenine adducts, but no bypass. There was no difference between the two sequence contexts or among any of the four adducts.

The Klenow fragment retains a 3′→5′ proofreading function, so it can degrade the primer to shorter lengths if the nucleotide supplied cannot be incorporated. With the hotspot template and 12-mer primer, thymine was preferentially incorporated opposite the adenine in the unmodified template, but some misincorporation occurred when nucleotides other than dTTP were present (Figure 6). We found that the (−) *syn* and (−) *anti* benzo[*a*]pyrene dihydrodiol epoxide adducts, which have the R configuration at position 10 of the opened epoxide, allowed incorporation of adenine or thymine across from the adduct prior to polymerase arrest. The adducts from the (+)-isomers, which have the S configuration at position 10, did not allow incorporation of any nucleotide across from the adduct. Thus, structural differences between the adducts with R versus S stereochemistry at position 10 led to different processing by the Klenow fragment. With the coldspot template, the adducts from the (−)-isomers responded as they did in the hotspot template, i.e., adenine or thymine was incorporated. The adducts from the (+)-isomers, however, also showed some misincorporation of adenine and a small amount of incorporation of thymine in this sequence context, whereas no incorporation occurred with the hotspot template.

Using HIV-1 reverse transcriptase, the hotspot template, and an 11-mer primer, we found that all four adducts allowed correct incorporation at position 12 prior to polymerase arrest with adducts from the (+)-isomers, whereas a base was incorporated across from the adducts derived from the (−)-isomers. Using the 12-mer primer, we could determine the base incorporated across from the adducts derived from the (−)-isomers (Figure 7). We found that the correct base, thymine, was incorporated and that the polymerase was blocked from further elongation. Thus, like the Klenow fragment, HIV-1 reverse transcriptase could distinguish between adenine adducts from the (+)- and (−)-dihydrodiol epoxide configurational isomers. However, HIV-1 reverse transcriptase incorporated only thymine opposite the adducts, whereas the Klenow fragment incorporated either adenine or thymine.

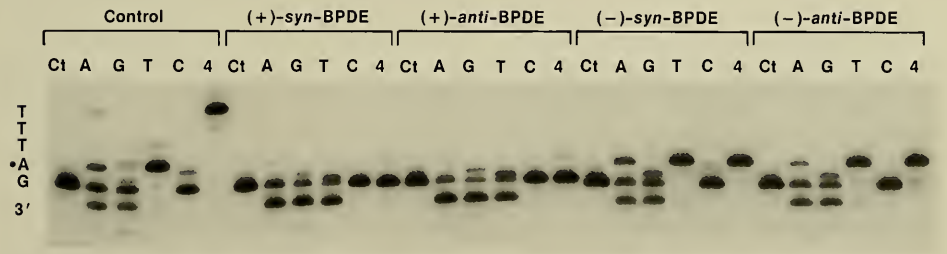


FIGURE 6. Effect of Klenow fragment on the hotspot template with the 12-mer primer. Each template-primer [unadducted DNA (control) or the four adducted templates] was incubated with individual nucleotides (A, G, T, C) or all four nucleotides (4) at a final concentration of 20 μM with Klenow fragment for 5 minutes at 25°C, then subjected to electrophoresis alongside a sample lacking polymerase (Ct). The template sequence is to the left of the gel, with the bottom nucleotide corresponding to the base that pairs with the primer terminus. The marked A denotes the adduct position.

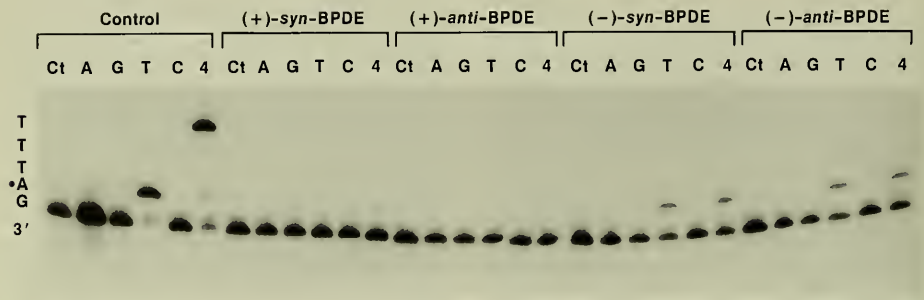


FIGURE 7. Effect of HIV-1 reverse transcriptase on the hotspot template with the 12-mer primer. The DNA samples were incubated with HIV-1 reverse transcriptase and the indicated nucleotides (at a final concentration of 100 μM) for 2 hours at 37°C. The lanes are labeled as described for Figure 6.

Overall, we have observed that polymerase action on these adducts depends on the sequence context of a given adduct, the adduct stereochemistry (i.e., R or S at C-10), and the choice of polymerase. A clear understanding of all these influences on the processing of adducts by polymerases should help to clarify the mechanisms through which these hydrocarbon adducts lead to deleterious biological effects.

Carcinogen-Modified Nucleic Acid Chemistry Section

Robert C. Moschel

Ronald E. Bishop
Mi-Young Chae
Mark G. McDougall
Victor Mikhailenko
Gary T. Pauly

Carcinogen damage to cellular DNA is considered a key step in the initiation of chemical carcinogenesis. The major goals of the Carcinogen-Modified Nucleic Acid Chemistry Section are to understand the chemical and biological consequences of this damage and to understand the role played by cellular repair mechanisms in modulating these effects. Because carcinogens react with a variety of sites on DNA bases and with several bases simultaneously, it is difficult to evaluate how any one carcinogen adduct alters the structure or coding properties of the damaged DNA. Therefore, our section exploits recent developments in synthetic organic chemistry and molecular biological techniques to prepare defined carcinogen-modified oligodeoxyribonucleotides (oligonucleotides) bearing specific carcinogen-modified bases at predetermined sequence locations. Using these substrates, we examine how individual adducts disrupt the structure and stability of DNA and, following insertion of modified substrates into plasmids, we examine the effects of adduct structure and sequence location on mutagenic potency. Recent experiments on DNA repair of our specifically modified substrates led to our discovery of a group of modified guanines capable of inactivating the human DNA repair protein O^6-alkylguanine-DNA alkyltransferase (AGT). These inactivating agents show great promise for use as adjuvants in chemotherapy for human tumors. Therefore, our section is developing additional potent and tumor-specific inactivating compounds for use as chemotherapeutic adjuvants.

During this report period, we examined the role of AGT in modulating mutagenesis by O^6-substituted guanines in codon 12 of the H-*ras* gene, we evaluated 13 new substituted O^6-benzylguanine derivatives as AGT inactivators to extend our studies of how substitution affects AGT-inactivating activity, and, in collaboration with Henry Friedman (Duke University Medical Center) and Clifford Schold (University of Texas, Southwestern Medical Center), we tested O^6-benzylguanine as an adjuvant for 1,3-*bis*(2-chloroethyl)-1-nitrosourea (BCNU) therapy for human medulloblastoma and glioblastoma multiforme tumors carried as xenografts in nude mice. The results of these studies are summarized below.

The Role of AGT in Protecting Rat4 Cells Against the Mutagenic Effects of O^6-Substituted Guanines Incorporated in Codon 12 of the H-*ras* Gene

In our earlier studies using an H-*ras* cassette vector (see 1989 Annual Report), we demonstrated that O^6-methyl- and O^6-benzylguanine residues could induce transformation of Rat4 cells when these adducts were positioned at either the first or second position of codon 12 in the rat H-*ras* coding sequence. However, since the levels of transformation were rather low, the data suggested that the major fate of these modified guanines was repair.

A major route for repair of O^6-alkylguanine residues in DNA is through reaction with AGT. In collaboration with Anthony Pegg (Pennsylvania State University College of Medicine) and M. Eileen Dolan (University of Chicago Medical Center), we recently showed that O^6-benzylguanine can serve as an alternative substrate for the protein and can deplete cells, tissues, and tumors of AGT activity.

Therefore, we compared the transforming potency of H-*ras* coding sequences containing O^6-methyl-(m^6G), O^6-ethyl-(e^6G), or O^6-benzylguanine (b^6G) residues, in place of the normal guanines of codon 12, in normal Rat4 cells and in cells grown in the presence of O^6-benzylguanine to completely deplete AGT. These data allowed us to evaluate the role of AGT in protection against the transforming potency of O^6-substituted guanines as a function of both their site of incorporation within codon 12 and the size of the substituent group attached.

Figure 1 shows the important structural features of the cassette plasmid used in these studies, together with a diagram for how carcinogen-modified DNA segments are inserted into these vectors. In the vector (Figure 1A), the

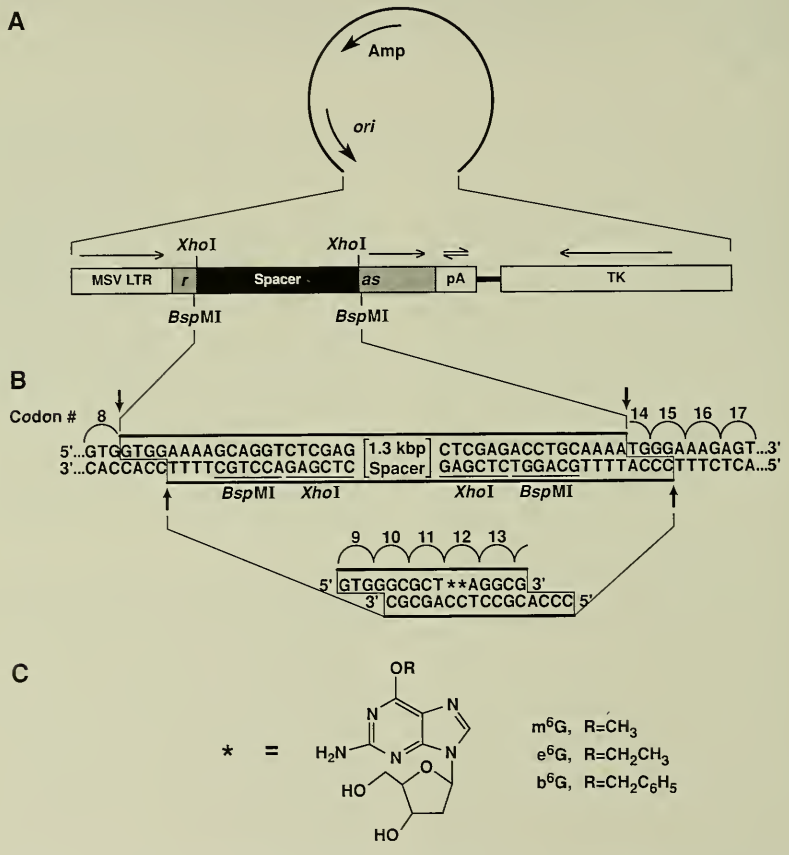


FIGURE 1. Construction of plasmids containing an O^6-alkylguanine-modified H-*ras* gene. The plasmid vector (A) was digested with *Bsp*MI (B). The major fragment liberated as a result of digestion was ligated to the synthetic oligodeoxyribonucleotide duplexes (C). The resulting H-*ras* coding sequences contained either an O^6-methylguanine (m^6G), O^6-ethylguanine (e^6G), or O^6-benzylguanine (b^6G) at the first, second, or both the first and second positions of codon 12.

H-*ras* sequences from codon 8 to the second base of codon 14 were separated by a 1.3-kilobase *Xho*I DNA "spacer" flanked by two outward-facing recognition sequences for the *Bsp*MI endonuclease. After digestion with *Bsp*MI, the indicated spacer and flanking sequences (Figure 1B) were removed and the major fragment resulting from *Bsp*MI digestion was purified and ligated to a synthetic oligodeoxyribonucleotide duplex (Figure 1C). The resulting DNA was transfected into cells grown in normal medium or medium containing O^6-benzylguanine to eliminate AGT activity.

The percentage of thymidine kinase-positive (TK^+) colonies transformed by H-*ras* genes carrying the O^6-substituted guanines is summarized in Figure 2. These data are normalized to the percentage of transformed colonies produced by transfection of a plasmid containing a mutated H-*ras* gene with the codon 12 sequence GAA. This latter percentage was in the range of 40% to 56% in each experiment and was set to 100%. The percentage of transformed colonies produced by transfection of plasmids containing the normal GGA codon 12 sequence was less than 0.2%. Figure 2 illustrates that the overall order of transforming potency for the modified guanines was $m^6G > e^6G > b^6G$, regardless of the AGT status of the cells. In AGT-competent cells, transformation due to O^6-methylguanine was unaffected by the site of incorporation within codon 12, in agreement with our earlier observations. In contrast, O^6-ethyl- and O^6-benzylguanine induced somewhat higher percentages of transformed colonies when located at the first position of codon 12 rather than at the second position. The results of transfection into AGT-depleted cells were somewhat different (Figure 2). In these cells, an increase in the percentage of transformed colonies was observed in all cases, although the percentage was still low. However, for the O^6-methylguanine case, the enhancement in transformation was greater when the modified base was incorporated at the first position of codon 12 rather than at the second position. Only vectors harboring adjacent m^6G residues induced substantially greater percentages of transformed colonies in AGT-depleted cells than did their singly modified counterparts.

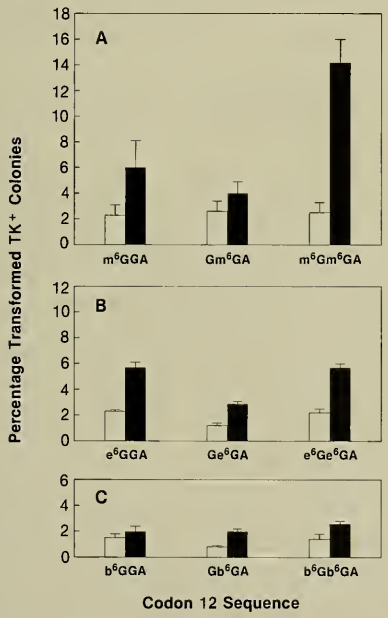


FIGURE 2. Percentage of transformed colonies resulting from transfection of vectors containing O^6-methylguanine residues (A), O^6-ethylguanine residues (B), or O^6-benzylguanine residues (C) in codon 12 of the H-*ras* gene in normal Rat4 cells (open bars) or in cells depleted of AGT (solid bars). Percentages of transformed TK^+ colonies are shown with respect to the site of incorporation within codon 12.

The Percentage of Mutations Induced by O^6-Methylguanine Residues in H-*ras* DNA

To determine the percentage of mutated H-*ras* in cells derived from exposure to O^6-methylguanine-containing plasmids, we used the *Gsu*I endonuclease, whose recognition site [CTGGAG(16/14)] encompasses codon 12, to analyze the amount of mutant DNA by restriction fragment length polymorphism. A 260-base pair DNA segment containing codon 12 was amplified using the polymerase chain reaction and radiolabeled nucleotide precursors. *Gsu*I cleaved up to 98.5% of the amplified DNA that contained a normal codon 12 GGA, which produced fragments of 168 and 92 base pairs. Amplified DNA was not cut by the enzyme if codon 12 was mutated (e.g., GAA). The percentage of total radioactivity associated with the uncut 260-base pair fragment is shown in Figure 3 as a function of both the site of incorporation of the modified base and

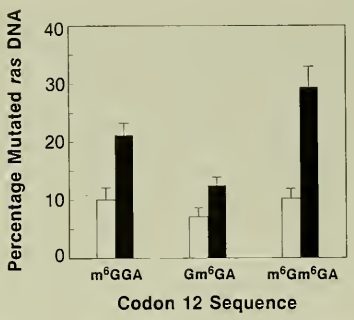


FIGURE 3. Percentage of mutated H-*ras* DNA in normal Rat4 cells (open bars) and cells depleted of AGT (solid bars) following transfection with O^6-methylguanine-containing vectors.

the AGT status of the transfected cells. In AGT-competent cells, 7% to 10% of the amplified H-*ras* DNA was mutated regardless of the guanine replaced by an O^6-methylguanine. In contrast, in AGT-depleted cells, the percentage of mutated DNA derived from the modified plasmids increased, but again the increase was greatest when transfectant DNA had an O^6-methylguanine residue at the first position of codon 12 rather than at the second. These data, together with the transformation data presented in Figure 2, indicate that AGT repairs an O^6-methylguanine residue somewhat more efficiently at the first position of codon 12 than at the second position.

Our data indicate that the contribution of AGT to the repair of O^6-substituted guanine damage in Rat4 cells decreases over the series methyl > ethyl > benzyl, i.e., the series of increasing size of the O^6-substituent. Additionally, for the case of O^6-methylguanine-containing plasmids, AGT appears to repair the adduct at the first position of codon 12 more rapidly than at the second position. Both these observations are consistent with previous in vitro observations. Our data for the fraction of mutated DNA produced by O^6-methylguanine-containing plasmids in AGT-depleted cells are very similar to the data presented by Ellison et al. (Proc Natl Acad Sci USA 86:8620, 1989), who measured the fraction of mutated DNA produced by m^6G-containing vectors in Mex^- Chinese hamster ovary cells. Significantly, both studies indicated that the fraction of mutated DNA, even in the absence of AGT, is far from 100% and suggest that significant repair of this base must occur by mechanisms other than those involving AGT. Excision repair mechanisms are known to be operative for bulkier adducts such as O^6-ethylguanine and perhaps for O^6-methylguanine as well, but rates for this process would have to exceed those for AGT-mediated repair in order to explain our results and those of Ellison et al. It must be remembered that site-specific mutagenesis techniques introduce very low levels of modified guanines into the DNA of transfected cells. These levels are several orders of magnitude lower than are usually produced when cells are exposed to alkylating *N*-nitroso compounds. It is not clear whether cellular repair mechanisms for dealing with very low levels of O^6-substituted guanine damage differ from those that are operative when cells are confronted with the much higher levels of damage produced by exogenous alkylating agents.

Activity of Substituted O^6-Benzylguanine Derivatives as AGT-Inactivating Agents

To extend our studies of how the size and position of substituents on O^6-benzylguanine alter its effectiveness as an AGT-inactivating agent, we prepared 13 new derivatives of this compound. These experiments provide information on where groups may be attached to O^6-benzylguanine without significantly lowering its AGT-inactivating activity. These sites can then be used for modifications that may alter the entry of O^6-benzylguanine into cells or direct it to tumor cells in preference to normal cells of a host.

The 13 compounds prepared for these tests are illustrated in Figure 4. Derivatives that are substituted on the benzene ring of the benzyl group [i.e., O^6-(*p*-bromobenzyl)-, O^6-(*p*-isopropylbenzyl)-, O^6-(3,5-dimethylbenzyl)-, O^6-(*p*-*n*-butylbenzyl)-, and O^6-(2-pyridylmethyl)-guanine] were prepared by reacting the respective sodium benzyloxide with 2-amino-6-chloropurine. O^6-Benzyl-9-methyl- and O^6-benzyl-7-methylguanine were isolated from reactions between the anion of O^6-benzylguanine and methyl iodide in *N,N*-dimethylformamide (DMF). O^6-Benzyl-9-acetylguanine and N^2-acetyl-O^6-benzylguanine were prepared by reacting O^6-benzylguanine with acetic anhydride in pyridine and toluene, respectively. The 9-pivaloyloxymethyl derivative of O^6-benzylguanine was isolated from the reaction of O^6-benzylguanine anion with

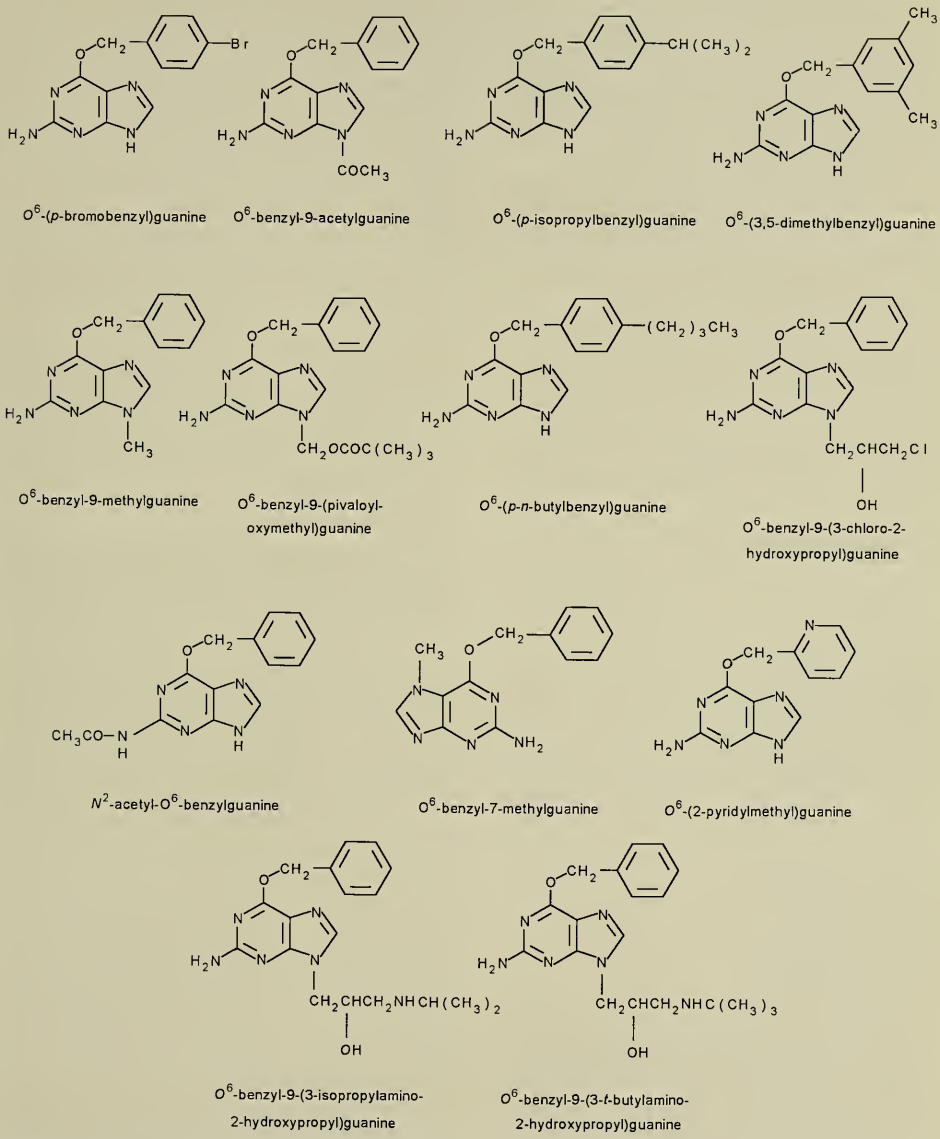


FIGURE 4. Structure of modified O^6-benzylguanine derivatives tested as inactivators of human AGT from HT29 colon tumor cells.

chloromethyl pivalate in DMF. O^6-Benzyl-9-(3-chloro-2-hydroxypropyl)guanine was prepared by reaction of O^6-benzylguanine with neat epichlorohydrin. Reaction of this latter material with an excess of isopropylamine or *t*-butylamine in dioxane produced O^6-benzyl-9-(3-isopropylamino-2-hydroxypropyl)- and O^6-benzyl-9-(3-*t*-butylamino-2-hydroxypropyl)guanine, respectively.

These compounds were furnished to Drs. Dolan and Pegg, who determined the ability of the compounds to inactivate the human AGT protein in cell-free extracts from HT29 cells and in intact HT29 cells. The results of these experiments are presented in Table 1.

As indicated, O^6-(*p*-bromobenzyl)guanine showed the highest AGT-inactivating activity. The 9-acetyl derivative of O^6-benzylguanine (O^6-benzyl-9-acetylguanine) (Figure 4) was the next most active compound; however, this material hydrolyzed readily in aqueous medium to liberate O^6-benzylguanine, which probably contributed to the higher than expected activity exhibited by this analog.

TABLE 1. AGT-depleting activity of O^6-benzylguanine derivatives

Compound	ED_{50} (μM)* In HT29 cell-free extract	In HT29 cells
O^6-(*p*-bromobenzyl)guanine	0.3	0.09
O^6-benzyl-9-acetylguanine	0.4†	0.14†
O^6-(*p*-isopropylbenzyl)guanine	0.5	0.6
O^6-(3,5-dimethylbenzyl)guanine	1.0	0.4
O^6-benzyl-9-methylguanine	2.6	0.4
O^6-benzyl-9-(pivaloyloxymethyl)guanine	3.1	0.3‡
O^6-(*p-n*-butylbenzyl)guanine	4.5	1.0
O^6-benzyl-9-(3-chloro-2-hydroxypropyl)-guanine	18	2
N^2-acetyl-O^6-benzylguanine	24	2
O^6-benzyl-7-methylguanine	52	17
O^6-(2-pyridylmethyl)guanine	58	16
O^6-benzyl-9-(3-isopropylamino-2-hydroxypropyl)guanine	106	23
O^6-benzyl-9-(3-*t*-butylamino-2-hydroxypropyl)guanine	106	26

*The effective dose required to produce 50% inactivation in cell-free extracts upon incubation for 30 minutes or in cells upon incubation for 4 hours. The corresponding values for O^6-benzylguanine are 0.2 and 0.05, respectively.

†Converted to O^6-benzylguanine by hydrolysis in vitro and in cultures.

‡Converted to O^6-benzylguanine by esterases in cell cultures. The compound is stable in vitro.

Tests with compounds bearing substituents around the benzene ring of the benzyl group showed a modest decrease in activity in the order O^6-(*p*-isopropylbenzyl)- > O^6-(3,5-dimethylbenzyl)- > O^6-(*p-n*-butylbenzyl)guanine. In contrast, increasing the steric bulk of substituents at the 9-position of O^6-benzylguanine decreased activity more significantly. Activity as a function of the 9-substituent decreased in the order 9-methyl (i.e., as in O^6-benzyl-9-methylguanine) > 9-pivaloyloxymethyl > 9-(3-chloro-2-hydroxypropyl) > 9-(3-isopropylamino-2-hydroxypropyl) ≥ 9-(3-*t*-butylamino-2-hydroxypropyl). Of the remaining compounds tested, the N^2-acetyl analog of O^6-benzylguanine (i.e., N^2-acetyl-O^6-benzylguanine), O^6-benzyl-7-methylguanine, and O^6-(2-pyridylmethyl)guanine exhibited intermediate activity. We conclude from these data that for AGT-inactivating activity, substituent changes around the phenyl ring of the benzyl group and at the 9-position of the purine are fairly well tolerated, although activity can vary dramatically with structural changes at the latter position.

Enhancement of BCNU Chemotherapy Against Medulloblastoma and Glioblastoma Multiforme by O^6-Benzylguanine

We recently completed collaborative studies with the laboratories of Drs. Friedman and Schold to determine whether O^6-benzylguanine in combination with BCNU would be effective in treatment of human medulloblastoma and glioblastoma multi-forme. These types of tumors in children are generally resistant to chloroethylating nitrosourea therapy or develop resistance to such therapy, which results in tumor progression and death. Fewer than 20% of children with these tumors survive more than three years following diagnosis. New therapeutic strategies against these tumors are clearly necessary. Since we showed previously that resistance to chloroethylating agents can be mediated by AGT, we compared the response of human medulloblastoma (D341 Med) and glioblastoma multiforme (D-456 MG) xenografts to treatment with BCNU alone or BCNU after preliminary exposure to O^6-benzylguanine to deplete AGT activity.

Table 2 shows the response of these two xenografts grown subcutaneously in athymic BALB/c mice to BCNU alone or O^6-benzylguanine in combination with BCNU.

TABLE 2. Response of human medulloblastoma (D341 Med) and glioblastoma multiforme (D-456 MG) xenografts to BCNU and O^6-benzylguanine*

	D341 Med		D-456 MG	
Treatment	T-C†	Regressions‡	T-C	Regressions
BCNU (13 mg/kg)	1.0	0/10	1.5	1/10
O^6-benzylguanine (80 mg/kg)	1.2	0/8	1.0	0/10
BCNU + O^6-benzylguanine	12.2	8/10	26.5	10/10

*Data of H.S. Friedman et al. (Duke University Medical Center).
†Difference between the median time (in days) for treated (T) and control (C) animal tumors to reach five times their tumor volume at time of treatment.
‡Defined as tumor volumes that decreased over two successive measurements at 3- to 4-day intervals.

While BCNU or O^6-benzylguanine alone had little if any effect on tumor growth, the combination of O^6-benzylguanine administered prior to BCNU resulted in prolonged growth delays and tumor regressions in almost all animals bearing the xenografts.

Figure 5 shows the survival time of BALB/c mice bearing intracranial xenografts of the human medulloblastoma D341 Med following treatment with drug vehicle (control), BCNU alone, or the combination of O^6-benzylguanine 1 hour prior to BCNU. Only the latter combination produced a significantly increased animal survival. These very promising subcutaneous and intracranial xenograft results offer additional strong support for tests of this adjuvant approach to nitrosourea therapy in humans. Preclinical toxicological evaluation of the combination of O^6-benzylguanine and BCNU will begin in the very near future. If satisfactory, Phase I clinical trials of this adjuvant approach in humans should begin shortly thereafter.

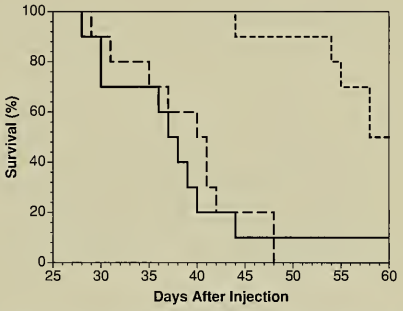


FIGURE 5. Survival of athymic BALB/c mice bearing intracranial D341 Med xenografts after treatment with O^6-benzylguanine and BCNU. The dose of O^6-benzylguanine was 300 mg/m^2. The dose of BCNU was 38 mg/m^2. Animals received either drug vehicle (——), BCNU alone (— —), or O^6-benzylguanine followed in 1 hour by BCNU (-----). Data of Schold et al. (Cancer Chemother Pharmacol, in press).

Laboratory of Chromosome Biology

Stuart J. Austin

Donald L. Court

The Laboratory of Chromosome Biology carries out basic studies of the replication, maintenance, and expression of genetic material. We primarily use the simplest possible prokaryotic models in our attempt to understand, at the molecular level, some of the most fundamental processes of living cells. *Escherichia coli* offers powerful and rapid genetic tools for the analysis of these processes. We are confident that expertise gained by this approach will then be applicable to analogous mechanisms in higher cells.

Knowledge of the pathways that govern the development of tumor cells has advanced greatly in the last few years, fueled in part by technologies and principles derived from prokaryotic genetics. However, the molecular mechanisms by which these pathways are regulated and ultimately act in the cell cycle are far from fully understood. Those transcription factors, proto-oncogenes, and tumor suppressor genes that interact directly with DNA do so in a highly specific manner in the normal cell. These specificities, we believe, may often be determined by complex interactions involving multiple protein components and altered DNA topology in much the same way as the key regulatory events that we study in our model prokaryotic systems. We believe that fundamental processes have a basic uniformity in nature, and that the study of simple prokaryotic models will continue to make major contributions to understanding them. By focusing on the most fundamental aspects of chromosome biology—the replication and maintenance of the genetic material and the general mechanisms by which this genetic material is expressed—we seek to maximize these contributions and to create unique opportunities for the productive exchange of ideas and technologies.

Cell Cycle Regulation Section

Stuart J. Austin

Ann Abeles
Therese Brendler
Michael Davis
Finbarr Hayes
Lyndsay Radnedge
Lucretia Reaves
Brenda Youngren-Grimes

The long-term goal of the Cell Cycle Regulation Section is to understand the function and interrelations of DNA replication, cell division, and chromosome segregation at the molecular level. Our initial aim is to understand how two fundamental events in this process occur: the controlled replication of the genetic material, and the equitable distribution of chromosomes to daughter cells. Our model systems consist of extrachromosomal elements, the low-copy-number plasmids P1 and P7, that mimic the activity of host chromosomes but are more amenable to genetic analysis because they are dispensable for cell viability.

Isolation and Specificity of the P7 Plasmid Centromere Analog

Low-copy-number plasmids are responsible for many economically important phenomena, such as drug resistance and bacterial pathogenicity. Many naturally occurring bacterial plasmids are stably maintained. This stability is particularly remarkable in the case of very low-copy-number plasmids, such as the *incY* plasmids P1 and P7 of *Escherichia coli*. These plasmids normally confer no selective advantage to their host. Yet, despite copy numbers as low as two per dividing cell, loss frequencies approximate one per million cell division events. In addition to efficient replication controls, this degree of stability requires a combination of several plasmid-encoded functions, including recombination systems that resolve plasmid multimers and activities that can kill the majority of plasmid-free cells that arise. However, the primary mechanism for ensuring the distribution of these plasmids involves active partition of copies to daughter cells in a fashion that, at least superficially, resembles the mitotic segregation of eukaryotic chromosomes.

The P1 partition system consists of two essential trans-acting genes, *parA* and *parB*, and a downstream cis-acting site, *parS*, which appears to be functionally analogous to a eukaryotic centromere. The *parB* gene encodes a site-specific DNA-binding protein (ParB) that has been postulated to recognize a set of heptamer repeats in *parS* (Figure 1). The host protein, integration host factor (IHF), binds cooperatively with ParB at *parS* (Figure 1), where IHF induces DNA bending in vitro (Funnell, J Biol Chem 266:14328, 1991). Although both the ParA and ParB proteins are essential for partition, no direct binding of ParA to *parS* has been detected. We have shown that ParA is a member of a diverse group of ATPases that function in plasmid maintenance. As the ParA ATPase is strongly stimulated in the presence of ParB, it is possible that ParA constitutes a karyokinetic motor that moves the ParB-plasmid complex during partition.

The active partition system of the P7 prophage plasmid is organized similarly and is homologous to that of P1, indicating that these regions are derived from a common progenitor. However, the systems have diverged sufficiently so that they demonstrate species specificity with respect to partition-site activities. Although the relative positions of the protein binding motifs in the P1 and P7 *parS* sites are conserved and the sites are 57% identical over the 94-base pair sequences illustrated in Figure 1, the P1 *parS* site fails to work with the P7 proteins, and vice versa (Table 1). Presumably, this specificity is due to key macromolecular

FIGURE 1. Alignments of the P1 and P7 *parS* regions. The smaller open boxes indicate the positions of the heptamer repeat sequences recognized in vitro by ParB. IHF motifs are marked by large boxes. The approximate limits of the P1 minimal *parS* sequence and the precise limits of the P7 *parS* sequence are shown by the bent arrows above and below the maps.

interactions in which the individual components of the P1 and P7 systems are not interchangeable. By pinpointing the elements that determine this specificity, we should be able to define crucial steps in the active partition pathway.

The Functional P1 and P7 Partition Sites. Successful partition-site assays utilize conditional vector systems that can be maintained at either high or low copy number. An assay suitable for the comparative analysis of both *parS* sites was developed. Potential partition sites are cloned into a mutant P1 miniplasmid that has a moderate copy number because the copy-control element *incA* has been deleted. The sites are tested after recombinational pick-up onto a phage lambda vector that supplies the missing P1 *incA* locus. The resulting composite replicates at low copy number, and the efficacy of the cloned *par* site can be estimated by measuring the maintenance stability of the recombinant when the appropriate Par proteins are supplied in trans from a compatible plasmid.

TABLE 1. Partition specificities of the P1 and P7 *parS* regions

		Percentage of retention in 25 generations (Par proteins supplied in trans)		
Plasmid	*parS* sequence	None	P1 ParA, B	P7 ParA, B
pALA1626	None	<2	<2	<2
pALA1775	P1 94 base pairs	<2	83 ± 6	<2
pALA1716	P1 88 base pairs	<2	90 ± 1	NT
pALA1791	P7 94 base pairs	<2	<2	75 ± 10
pALA1659	P7 78 base pairs	<2	NT	74 ± 3

All values are the averages of at least three individual experiments ± 1 SD. The indicated plasmid was incorporated into the lambda pick-up vector λ-P1:5RΔ1005 and tested for retention during 25 generations of unselected growth in the presence of pBR322 (which supplies no partition proteins), or pALA271 or pALA1023, which supply the P1 and P7 ParA and ParB proteins, respectively. NT, not tested.

A fragment encompassing the entire P1 *parS* region (94 base pairs, Figure 1) confers stability in response to the P1 Par proteins. The minimal functional P1 *parS* site in this assay requires approximately 88 base pairs of P1 sequence (base pairs 7 to 94), as shown in Figure 1. Shorter sequences, including the 22-base pair *parS* core site that has some activity in an alternative assay, are nonfunctional in this assay.

Figure 1 shows the region immediately downstream of P7 *parB* aligned with the 94-base pair P1 *parS* sequence. This alignment reveals a related group of identically spaced protein-binding motifs in the P7 *parS* region. To investigate whether this P7 region also functions as a partition locus, the equivalent 94-base pair fragment spanning the region (Figure 1) was tested. The fragment promoted stability of the lambda composite when the P7 Par proteins were supplied in trans (Table 1). Partition tests using these fragments were species specific: the partition-site activity of the P7 fragment was supported when P7 proteins were provided in trans and not when P1 proteins were supplied, while the derivative containing the P1 94-base pair *parS* fragment was dependent on the P1 proteins (Table 1). We determined the exact extent and location of the P7 partition-site sequences, using constructs that lacked various lengths of sequence at the right or left ends of the 94-base pair P7 fragment. The 78-base pair sequence (base pairs 1 to 78, Figure 1) was thus defined as being the minimal P7 partition site (Table 1).

Site-Directed Mutational Analysis of the P7 Minimal *parS* Region. We examined the roles of selected sequences in the P7 minimal *parS* region by cloning synthetic oligonucleotides with multiple mutations in two of the three ParB heptamer repeat boxes, in the IHF-binding site, and in the region adjacent to the leftmost boundary of the minimal *parS* site (Figure 2). Mutations in the IHF box (pALA1759),

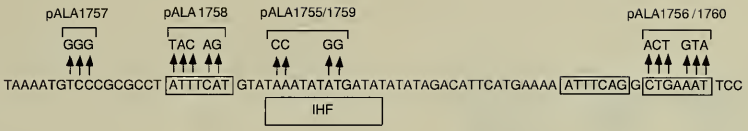


FIGURE 2. Mutational analysis of the P7 minimal *parS* region. Heptamer repeats recognized by ParB are boxed. The position of the IHF-binding motif is indicated. Substitution mutations generated in the sequence are shown above the *parS* sequence. The plasmid numbers for clones containing each set of mutations are shown above the substitutions.

the region close to the leftmost limit of the minimal site (pALA1757), and the rightmost ParB heptamer repeat (pALA1760) fully abolished partition activity, whereas a plasmid containing the substitutions in the leftmost ParB box (pALA1758) was retained at approximately 50% the level of a plasmid containing the wild-type sequence (pALA1659; Table 2).

Protein Binding to Wild-Type and Mutated P7 Minimal Partition-Site Sequences. The in vitro binding patterns of purified IHF and P7 ParB proteins to wild-type and mutated P7 minimal partition sites were examined by the DNaseI protection (footprinting) technique. Sequences at the rightmost boundary of the wild-type site were protected from DNaseI digestion by ParB binding (Figure 3, lower region of gel). This region includes two inversely oriented copies of the proposed ParB heptamer recognition element, with the sequences ATTTCAG and CTGAAAT (Figure 2). ParB also weakly protected a region that includes the leftmost ParB heptamer ATTTCAT (Figure 2) and extends somewhat beyond the left boundary of the partition site (Figure 3, upper region of gel). IHF protected a region that includes the 13-base pair putative IHF-binding motif and additional bases, mostly to the right of the motif. ParB and IHF together gave a pattern similar to the sum of the separate effects, but the protection was more complete. Protection from the left boundary through the rightmost heptamer was then clearly evident.

Although the rightmost ParB heptamer motif was rather poorly cut by DNaseI, inspection of Figure 3 (and of overexposed versions not shown) shows that ParB protects it and that this protection is complete in the presence of IHF. The introduction of multiple mutations into this box (pALA1756) altered and improved the basic DNaseI cutting pattern of the region, making the protection experiments easier to interpret. ParB binding to this mutated site was eliminated, but binding to the adjacent box persisted. Thus, the heptamer motif is an individual ParB binding site, as we have previously speculated for the P1 system.

The set of mutations introduced into the putative IHF box (pALA1755; Figure 2) completely eliminated the recognition of this site by IHF protein (Figure 3), confirming the identity and location of the IHF site (Figures 1 and 2). The pattern of ParB binding to this mutant was unaltered compared to the pattern of ParB bound alone to the wild-type site (Figure 3).

TABLE 2. Partition properties of P7 minimal *parS* sites containing multiple mutations

Plasmid	*parS* region	Percentage of retention in 25 generations (Par proteins supplied in trans)	
		None	P7 ParA, B
pALA1626	None	<2	<2
pALA1659	Wild type	<2	74 ± 3
pALA1757	Left boundary mutant	<2	<2
pALA1758	Leftmost ParB box mutant	<2	33 ± 10
pALA1759	IHF motif mutant	<2	<2
pALA1760	Rightmost ParB box mutant	<2	<2

Partition-site tests were carried out as described in the legend to Table 1. Plasmid pALA1626 derivatives contained P7 base pairs 1 to 78 with or without multiple mutations in the indicated regions, as shown in Figure 2.

Exploitation of Partition Specificity. The partition systems of P1 and P7 show species specificity—P1 partition proteins fail to work with the P7 *parS* site, and vice versa. The cis-acting *parS* sites therefore contain sequences required for differential recognition and/or functional interaction with the partition proteins of P1 and P7. Figure 1 shows that the relative spacings of the ParB and

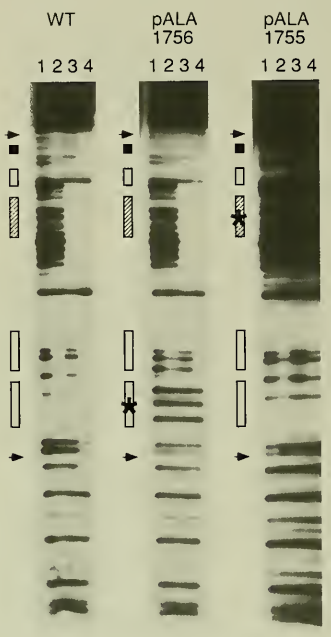


FIGURE 3. DNaseI footprinting patterns of wild-type and mutated P7 minimal partition-site sequences with IHF and P7 ParB proteins. The P7 DNA consists of the 78-base pair partition site (P7 base pairs 1 to 78) inserted into pBR322. Arrows mark the boundaries of the P7 and vector bases. Open and hatched boxes indicate the positions of the putative ParB-binding heptamers and IHF-binding motif, respectively. Asterisks mark the boxes containing multiple mutations (see Figure 2). WT, wild-type sequence; pALA1756, multiple mutations introduced into the rightmost ParB box; pALA1755, multiple mutations introduced into the IHF motif. Lane 1, no protein added; lane 2, 6.0 μg/ml P7 ParB; lane 3, 0.2 μg/ml IHF; lane 4, 2.0 μg/ml ParB and 0.2 μg/ml IHF.

IHF protein binding motifs are conserved perfectly and that there is no clearly defined consensus difference between the P1 and P7 heptamer boxes. The overall form and configuration of these sites, when present in an active partition complex, are probably similar. How then is the specificity difference determined?

The extreme ends of the P7 *parS* site clearly contain essential information of unknown function. The deletion analysis described here demonstrated that 16 base pairs at the left end and 3 base pairs at the right are necessary for *par* activity but do not form part of any obvious ParB heptamer or IHF-binding box. These sequences differ radically from their P1 counterparts, suggesting that they play a role in species specificity. We are analyzing the role of these sequences by using synthetic hybrid *parS* sites in which blocks of P1 information have been substituted for P7 information and vice versa. We are also creating and testing hybrid ParA and ParB proteins to map the domains responsible for species-specific DNA-binding and protein-protein interactions. In this way, critical contacts between these molecular components could be defined, enabling us to develop a molecular model of the partition process.

Evidence for Two Levels of Control of P1*oriR* and Host *oriC* Replication Origins by DNA Adenine Methylation

The prophage of bacteriophage P1 is maintained in its *E. coli* host as an autonomous plasmid whose replication is under stringent control. The plasmid has multiple DNA adenine methylation (GATC) sites clustered within the origin, P1*oriR* (Figure 4). These sites are rare in bacterial plasmid origins, but they are found in the origin of the host bacterial chromosome, *oriC*. Wild-type P1 miniplasmids cannot be maintained in strains that lack DNA adenine methyltransferase activity (Dam$^-$ strains), suggesting a stringent requirement for adenine methylation for the origin to fire. However, origin methylation is not essential for *oriC*, as it is able to function in Dam$^-$ strains.

In the case of *oriC*, methylation is implicated in the timing of initiation within the cell cycle. Immediately following replication from the fully methylated *oriC* origin, the GATC sites are hemimethylated. The origin region containing these hemimethylated sites is sequestered by binding to the bacterial membrane (Ogden et al., Cell 54:127, 1988). This sequestration prevents a second round of replication until methylation is completed—an event that takes a considerable portion of the cell cycle for the origin GATC sites (Campbell and Kleckner, Cell 62:967, 1990).

Although P1*oriR* and *oriC* seem to respond to methylation differently, the similarity of organization of the origins and their methylation sites suggests some commonality of mechanism. We examined this possibility.

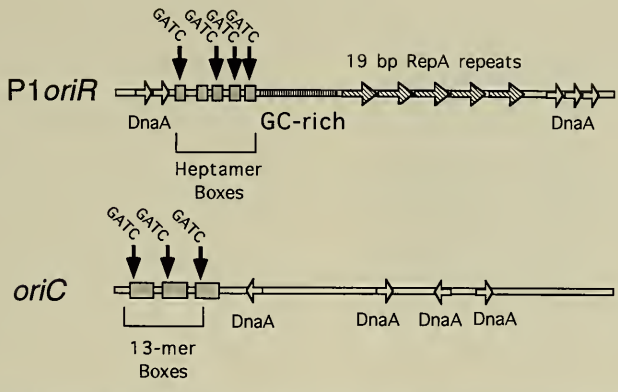


FIGURE 4. Physical maps of the P1 *oriR* and *E. coli oriC* replication origins. Open and hatched arrows mark the repeat sequences responsible for binding of the DnaA and P1 RepA proteins, respectively. The heptamer and 13-mer repeat sequences (stippled boxes) contain clustered GATC sequences that are substrates for DNA adenine methylase.

A High-Copy-Number P1 Miniplasmid Can Transform Dam⁻ Strains. A plasmid construct driven by the wild-type P1 plasmid replicon is unable to become established in Dam⁻ strains (pALA109, Table 3). However, the mini-P1 plasmid pSP102, which lacks the P1 copy-control element *incA*, did give some transformants. This plasmid retains all the elements essential for replication, including the origin with its multiple GATC sites. However, unlike wild-type mini-P1 plasmids that have copy numbers as low as 1 to 2 per cell, pSP102 is maintained at a moderate copy number (8 to 10) in wild-type hosts (Figure 5). In this respect, pSP102 resembles *oriC* plasmids that also have elevated copy numbers even though their replication machinery is derived from the *E. coli* chromosome (Figure 4).

Russell and Zinder (Cell 50:1071, 1987) have shown that *oriC*-driven plasmids completely lacking methyladenine are able to transform Dam⁻ strains relatively efficiently and can be propagated in the plasmid state. However, methylated *oriC* plasmid DNAs transformed very inefficiently. These plasmids were replicated once, but further replication of most of the copies was blocked at the hemimethylated stage. We carried out a similar experiment using the P1-derived plasmid pSP102. This plasmid transformed a wild-type host readily, and plasmid DNA recovered from these cells was fully methylated. Methylated DNA extracted from these cells transformed a Dam⁻ strain approximately 1000-fold less efficiently than an isogenic Dam⁺ strain (Table 3). A similar effect was evident with the *oriC* plasmid pOC15 (Table 3) in conformity with the original observation. The few pSP102 colonies that arose from the Dam⁻ host contained low levels of plasmid DNA when propagated under selection (Figure 5, lane C), and this DNA proved to be fully unmethylated when further analyzed.

TABLE 3. Frequency of transformation of Dam⁺ and Dam⁻ strains by P1 *oriR* and *oriC* miniplasmids

Transforming DNA		Strain transformed (plasmid-containing colonies/μg DNA)		
		Dam⁺	Dam⁻	Dam⁻/Dam⁺
pALA109	(Meth⁺)	7×10^4	<10	$<1 \times 10^{-4}$
pOC15	(Meth⁺)	2×10^5	2×10^2	1×10^{-3}
pSP102	(Meth⁺)	5×10^5	4×10^2	1×10^{-3}
pSP102	(Meth⁻)	8×10^5	1×10^6	1
pSP102	(Meth⁺)*	2×10^5	2×10^1	1×10^{-4}

The terms Meth⁺ and Meth⁻ indicate that the transforming DNA was prepared from Dam⁺ or Dam⁻ hosts, respectively.

*The DNA used for these transformations was recovered from the Meth⁺ transformants that had been transformed using the Meth⁻ DNA to show that the increased transformation frequencies found with the Meth⁻ DNA was not due to accumulation of mutations.

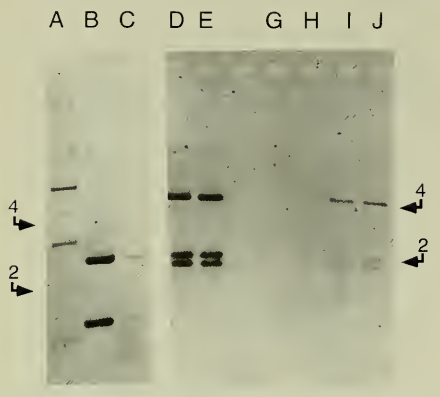


FIGURE 5. The copy numbers of the P1*oriR* and *oriC* plasmids are low in Dam⁻ strains. Relative plasmid DNA contents of cells carrying the P1*oriR* plasmid pSP102 (lanes B and C) and the *oriC* plasmid pOC15 (lanes D to J) are shown. The wild-type mini-P1 plasmid pALA109 (lane A), which is present at 1 to 2 copies per host chromosome, serves as a standard for quantitation. The lanes contain: A, pALA109(Meth⁺) digested with *Hin*dIII; B, pSP102 (Meth⁺) digested with *Eco*RI; C, pSP102(Meth⁻) digested with *Eco*RI; D and E, pOC15(Meth⁺) digested with *Hin*dIII; and G to J, pOC15(Meth⁻) digested with *Hin*dIII. The arrows marked 2 and 4 indicate DNA sizes in kilobases.

Unmethylated P1*oriR* DNA Transforms Dam⁻ Strains With Greatly Increased Efficiency. The small quantities of unmethylated mini-P1 DNA recovered from the Dam⁻ strain were used to retransform the Dam⁻ strain. This DNA was 1000-fold more efficient than fully methylated DNA at transforming the Dam⁻ strain (Table 3). The improved transforming ability was not due to spontaneous mutation of the DNA, as further passage through a Dam⁺ strain fully restored the properties of the derivatives of the unmethylated DNA pool to those of methylated DNA (Table 3). We conclude that unmethylated mini-P1 DNA is able to bypass a block to plasmid establishment that normally occurs when a methylated plasmid is introduced into a Dam⁻ strain—a property shared with *oriC* plasmids.

Mini-P1 Plasmid Copy Number Is Low and Maintenance Efficiency Is Poor in Dam⁻ Cells. The plasmid content of the Dam⁻ cells transformed with mini-P1 pSP102 is much less than that of the equivalent Dam⁺ strain (Figure 5) when grown under selection for plasmid retention. This low copy number was reproducible when we examined a number of independently transformed clones containing pSP102. As this plasmid lacks an active plasmid partitioning system, its maintenance stability should be roughly proportional to plasmid copy number. In the absence of selection, the plasmid was relatively stable in Dam⁺ cells, but was rapidly lost from Dam⁻ cells (Table 4).

Thus, although pSP102 can be established in a Dam⁻ strain without continuous methylation, replication is inefficient under these conditions, so that the average copy number of the plasmids is low and loss of the plasmid during cell division is high. We presume that the wild-type mini-P1 also replicates inefficiently in Dam⁻ strains and that the combination of this inefficiency and the stringent copy control imposed by the P1*incA* locus results in a plasmid that is so unstable that its establishment is difficult to demonstrate. Inefficient replication of pSP102 was shown to be due specifically to a lack of adenine methylation, as the deficiency was fully reversed when plasmids carrying the *E. coli* or phage T4 Dam methylases were introduced into these strains (data not shown). As the phage-encoded enzyme is very different in primary structure from its host counterpart, we conclude that Dam

TABLE 4. Plasmid maintenance in Dam⁺ and Dam⁻ strains

Strain	Retention of plasmid in 25 generations of unselected growth (%)	
	pSP102	pOC15
MC1061*dam*⁺	44	48
S1540*dam*⁻	<1	(100)*
S1540*dam*⁻/pDAM118	51	NT
S1540*dam*⁻/pSSH6	44	NT

Retention of the pSP102 or pOC15 was estimated by scoring retention of antibiotic resistance after approximately 25 generations of unselected growth at 30°C. Resident plasmids pDAM118 and pSSH6 supply the *E. coli* and T_4 Dam methylases, respectively. NT, not tested

*Colonies of the primary transductants on selective medium grow very poorly and contain many cells without the plasmid. However, cells that retained the plasmid antibiotic-resistance marker subsequently failed to lose it, even after extensive growth without selection. We conclude that the plasmid is very poorly maintained and that selection rapidly gives rise to clones in which the plasmid has integrated into the host chromosome.

methylase per se is required for efficient origin function rather than for interaction of the origin with the methylase protein itself.

The Copy Number and Maintenance Efficiency of an *oriC* Plasmid in a Dam⁻ Host Is Also Low. The similarity of the properties of mini-P1 and *oriC* plasmids on transforming Dam⁻ cells prompted us to examine whether *oriC* plasmids are also poorly maintained once established in Dam⁻ strains. We prepared plasmid DNA from several pOC15 Dam⁻ transformants (Table 3) grown under ampicillin selection. In contrast to Dam⁺ control transformants, these cells yielded little or no plasmid DNA, giving a maximum of eightfold less DNA than was recovered from the Dam⁺ clones propagated under the same conditions (Figure 5). Further analysis showed that these *oriC* plasmids were highly unstable but, when under selection, rapidly gave rise to clones that stably maintain the plasmid antibiotic-resistance marker (Table 4). We conclude that the *oriC* plasmid is replicated very inefficiently in the Dam⁻ strain, which leads to a much lower average copy number and very poor plasmid maintenance. These cells have a severe growth disadvantage when grown in the presence of ampicillin, and variants are selected in which a copy of the plasmid (or at least the ampicillin-resistance gene) is integrated into the chromosome.

The Methylation State of the Origin Affects Replication in Vitro. We investigated the activity of hemimethylated DNA in our in vitro replication system (Table 5). The template consists of the minimal P1 origin introduced into the *Eco*RI site of the vector f1h$_o$. This vector is a modified f1 phage that has no adenine methylation sites. Thus, any effects of adenine methylation on the template can be attributed directly to effects on the inserted P1 origin DNA. By demonstrating dependence on addition of the P1 initiator protein RepA, incorporation in each assay was shown to originate from the P1 origin.

The system gave good incorporation with fully methylated template DNA and eightfold less incorporation with fully unmethylated DNA templates (Table 5). When this unmethylated DNA was treated with Dam methylase in vitro prior to use in the system, template efficiency was restored (Table 5). Thus, efficient activity in the in vitro system is dependent on template methylation. As the only adenine-methylation sites present are in the P1 origin region, we conclude that methylation of the P1 origin is required for efficient initiation.

TABLE 5. In vitro replication from P1*oriR* requires methylated or hemimethylated DNA

Plasmid DNA used	Methylation in vitro*	pmol ^{3}H incorporated	
		No RepA	RepA
Vector (f1h$_o$)	No	11	11
f1P1 (Meth⁺)	No	10	174
f1P1 (Meth⁺)	Yes	8	152
f1P1 (Meth⁻)	No	12	39
f1P1 (Meth⁻)	Yes	8	210
f1P1 (Hemimethylated)†	No	10	110
f1P1 (Hemimethylated)	Yes	12	103

The f1P1 template has the P1 origin region (base pairs 373 to 606) cloned into the *Eco*RI site of the f1h$_o$ vector. Replication was measured as pmol of [^{3}H]dTTP incorporated into a TCA-precipitable fraction using the *E. coli* in vitro replication system. The reactions contained 2 mol added DnaA protein, 0.04 pmol DNA, and 6 pmol RepA when present.

*The DNA was methylated in vitro with purified Dam and S-adenosyl methionine for 30 minutes at 30°C. The RepA and FII extract were then added and replication was measured as before.

†Hemimethylated DNA was prepared from linear double-stranded methylated replicative form DNA and unmethylated circular phage DNA.

Hemimethylated, supercoiled DNA was constructed by annealing single-stranded, unmethylated viral f1P1 DNA to separated strands of the fully methylated DNA and ligating the single-strand breaks under conditions that yield covalently closed supertwisted circles. As shown in Table 5, the hemimethylated template was active in vitro. The incorporation achieved with this DNA was somewhat less than that with fully methylated template. This effect is probably due to some damage to the hemimethylated DNA, as shown by pretreating the hemimethylated DNA with Dam methylase. No further stimulation of template activity by this pretreatment was evident (Table 5), despite the fact that the DNA was then fully methylated on both strands (data not shown). Thus, at least one of the hemimethylated forms normally produced by replication (that with the methyl groups on the top strand of the conventional map, Figure 4) is fully competent for replication in vitro. Preliminary evidence suggests that the other hemimethylated form is also active (data not shown).

The Similar Properties of P1 *oriR* and Host *oriC* Plasmids Suggest That They Have Similar Replication Controls. Our results suggest that the methylation controls of the *oriC* and P1*oriR* plasmids are very similar. Plasmids driven by *oriC* are subject to a well-documented negative control system acting at the hemimethylated level. The results presented here strongly suggest that P1*oriR* is subject to a similar type of regulation. However, we detected an additional level of control. Both mini-P1 and *oriC* plasmids can be maintained in a Dam$^-$ strain, but the resulting replication is inefficient; the plasmid is unable to maintain its normal copy number and is frequently lost. We conclude that both P1*oriR* and *oriC* origins function poorly when unmethylated.

Like *oriC*, P1 *oriR* appears to be subject to a negative control in which hemimethylation causes a temporary block to reinitiation, presumably due to membrane sequestration. This control is not expected to operate in the in vitro system, because membranes are absent. Thus, we can explain why the hemimethylated P1 origin was found to be an efficient template in vitro. The inefficiency of replication of both types of origin in Dam$^-$ cells is, however, paralleled by the behavior of the plasmids in vitro; unmethylated templates of both plasmid types are severalfold less efficient than their methylated counterparts.

Although methylation appears to act on mini-P1 and *oriC* plasmids at two different levels, the two effects could formally be manifestations of the same mechanism. For example, if membrane sequestration brings the origin sequences into a cell compartment that is particularly favorable for subsequent initiation events, a complete lack of methylation might be disadvantageous. However, we favor models in which the two levels of control are due to independent mechanisms. For example, in addition to recognition and sequestration of the hemimethylated origin DNA by membrane sites, some factor might exist that is required for optimum origin efficiency and binds to the origin only in its fully methylated state. Our in vitro results with P1*oriR* bear on this question. The P1 in vitro system does not contain membranes and shows no discrimination against hemimethylated DNA. Yet, the system shows a strong discrimination against the unmethylated P1*oriR* origin. Thus, the unmethylated origin appears to be inherently inefficient, even when no hemimethylation block is operating. This ability of the in vitro system to uncouple the two types of effect caused by adenine methylation strongly suggests that they are due to separate mechanisms.

Molecular Control and Genetics Section

Donald L. Court

Teresa Baker
Stanley Brown
Nina Costantino
Asis Das
Santanu Dasgupta
Bradford Powell
Helen Wilson

The goal of the Molecular Control and Genetics Section is to understand the molecular interactions that occur during gene regulation. Primarily, we study the regulation of transcription termination and the regulation of translation initiation. Most of our studies, whether transcriptional or posttranscriptional, involve mechanisms of protein-RNA interactions that lead to changes in gene expression. Here, we will review three areas in which we have made progress during the past year. One involves transcription termination and, specifically, what is required in vivo for efficient Rho-dependent termination of RNA polymerase transcription. A second area involves one aspect of our studies on the RNaseIII operon. Specifically, we will review our isolation and characterization of mutants that suppress temperature-sensitive lethality caused by a defect in the *era* gene within the RNaseIII operon. The third area of study involves a novel protein display library of synthetic peptides on the surface of *Escherichia coli*, and the selection from this library of cells that bind to specific monoclonal antibodies against the reverse transcriptase (RT) of human immunodeficiency virus (HIV).

Transcriptional Control of Gene Expression

Transcriptional Termination and Antitermination. Transcription termination and antitermination are critical points of transcriptional control of gene expression common to both prokaryotes and eukaryotes. *E. coli* and phage lambda have the best-studied and most extensively characterized examples of transcription-termination control. There are two general types of terminators in *E. coli*: intrinsic terminators whose signals are recognized by RNA polymerase in a pure in vitro system, and Rho-dependent terminators, which require Rho protein in addition to RNA polymerase. Transcription elongation through both of these types of terminators can be regulated to mediate the extent of distal gene expression.

Transcription termination and antitermination are the major mechanisms by which gene expression is temporally controlled in phage lambda development. The first lambda protein expressed following infection is that from the *N* gene. The major function of the N protein is to positively control expression by interacting with the RNA polymerase transcription complex on lambda and causing polymerase to pass through normal transcription stop points. Both Rho-dependent and intrinsic termination are antiterminated by N.

Rho-Dependent Termination. As discussed in our previous Annual Reports, we have studied transcription termination as well as N-mediated antitermination at the Rho-dependent t_{R1} terminator of phage lambda. A 400-base pair DNA fragment from lambda containing *cro-nutR-*t_{R1}*-cII* was inserted into a plasmid such that the *lac* promoter is upstream of *cro* and the *galK* reporter gene is downstream of *cII*. Galactokinase levels from *galK* in the resulting plasmid (pMZ245) are an indication of the net effects of the *lac* promoter and the t_{R1} terminator. Termination at t_{R1} on the vector is Rho dependent as it is on lambda; when N function is provided in the cell, N binds to the *nutR* site on the RNA and prevents termination at t_{R1} (Figure 1).

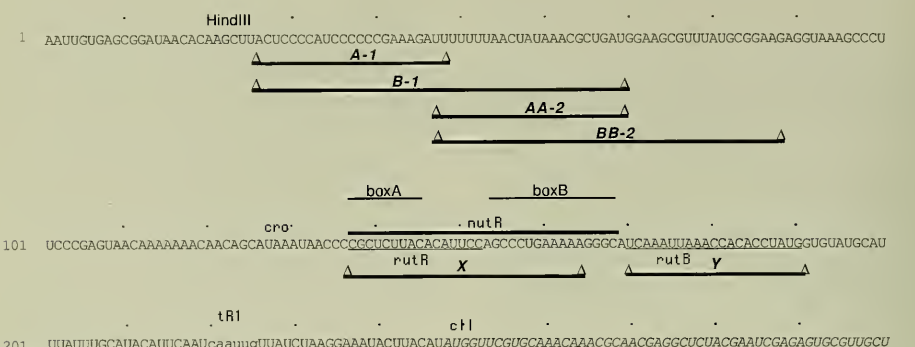


FIGURE 1. The t_{R1} transcript of pMZ245. The entire *cro*-t_{R1}-*cII* segment is shown. The *cro* gene is indicated at its UAA stop codon. The *cII* gene is indicated by italics. The Rho binding sites, *rutA* and *rutB*, are indicated. The overline represents the *nutR* region required for *N* recognition during antitermination of transcription. The *boxA* and *boxB* segments of *nutR* are shown. The site where Rho causes termination at t_{R1} in vitro is indicated by lowercase letters. The *Hin*dIII site is the position at which the *lac* promoter was placed upstream of the *cro* segment in pMZ245. The transcript shown starts at base 1 from the *lac* promoter of pMZ245. The horizontal bars with delta elements at each end represent the deletions shown in italics above each bar. Deletion letters followed by –1 indicate part of the set of nested deletions for experiment 1. Deletion letters followed by –2 indicate part of the set of nested deletions for experiment 2. The termination data for the full set of experiments in each set are shown in Figure 2. Deletion X removes the *rutA* site; deletion Y removes the *rutB* site.

In vitro studies have suggested that there are certain requirements for efficient Rho-mediated termination. For Rho to cause termination, the transcribed RNA length must be at least 80 nucleotides. Because Rho first binds to the nascent RNA and then interacts with the transcribing RNA polymerase, it requires the 80-nucleotide length of RNA for efficient binding. Rho is known to bind best to polyC RNA, as its RNA-dependent ATPase activity is very active with polyC and nearly inactive with polyA RNA. Therefore, in addition to the length of nascent RNA, the base distribution is also important. For efficient termination in vitro, cytidines must be distributed throughout the RNA, although no strict number or order of the cytidines along the RNA is known to be important. There are sites, called *rut*, upstream of t_{R1} that are important in vitro for tight binding and necessary for efficient termination. However, *rut* sites have not been found at other Rho-dependent terminators.

Our results (1992 Annual Report) support the in vitro finding that a minimal RNA length is required for termination and that cytidines are important for efficient termination. The frequency of termination at t_{R1} approaches 0 when RNA length between the start site of transcription and t_{R1} is less than 80 nucleotides and increases to greater than 0.95 when RNA length exceeds 240 nucleotides. In addition, we determined that an RNA segment that is rich in cytidines and located within 40 nucleotides of the 5′ start is critical for efficient termination (shown by deletion A-1 in Figure 1). To explain these and other in vivo results, we propose that Rho binds to the RNA transcript preferentially near the 5′ end and translocates along the RNA in a 5′ to 3′ direction, using its RNA-dependent ATPase activity. It is assumed that Rho must move faster along the RNA than RNA polymerase to catch the polymerase while it pauses at the t_{R1} terminator, where termination then occurs.

We have reexamined the requirement for Rho binding to RNA by comparing RNAs of various lengths in two sets of experiments (Figure 2). One set maintained a cytidine-rich RNA segment 40 bases from the 5′ end of the RNA, and the other set contained various sequences in place of the cytidine-rich segment. The results show that, when the first set is compared with the second, the cytidines enhance termination efficiency over the entire range of lengths tested, indicating that cytidines are a limiting factor for

efficient Rho termination and that their location near the 5′ end is important. When these 5′ cytidines are removed, termination is greatly reduced, unless another group of cytidines is moved near the 5′ end to replace the deleted cytidines. This is particularly apparent for the first two deletions, A-1 (199 nucleotides) and B-1 (179 nucleotides), of SET 1 in Figure 2. The results also show that the minimal RNA length for Rho to function is approximately 70 to 90 nucleotides, as determined in vitro (Figure 2).

We have also introduced base substitution and deletion mutations into the *rut* sites upstream of t_{R1} to test their requirement for termination in vivo. As described previously, deletions or substitutions of *rutA* have only minor effects on termination (1992 Annual Report). Similar results have been found for deletion or substitution of the *rutB* sequence (Table 1). These in vivo results indicate that *rutA* or *rutB* is neither sufficient nor required for efficient termination at t_{R1}, using the plasmid described here. On the other hand, the cytidine-rich region near the 5′ end is important for termination, as deletion of this region (A-1) increases readthrough of t_{R1} and galactokinase expression 15-fold (Table 1).

TABLE 1. Effects of *rutA* and *rutB* on termination at t_{R1} in deletion derivatives of pMZ245

Mutation*	Percentage of galactokinase†	Percentage of termination‡
None	3	97
A-1	45	55
AA-2	3	97
X	9	91
Y	7	93

*Deletion in pMZ245 that is described in Figure 1.

†Total transcription in the absence of the t_{R1} terminator yielded 2400 units of galactokinase in this experiment. Values shown are normalized to this value, which was set as 100%. Higher numbers indicate weaker terminator strength.

‡Transcription termination is 100% minus the percentage of galactokinase.

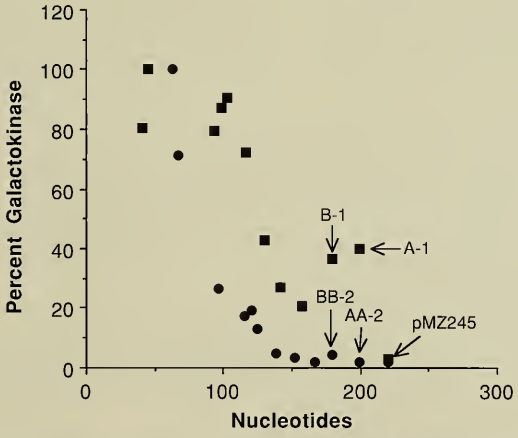


FIGURE 2. Nucleotide length analysis of Rho-dependent termination. Galactokinase values were determined from two sets of plasmid derivatives of pMZ245. Each set contained 12 plasmids, including pMZ245. The plasmids differed by deletions located between the promoter and terminator, thereby generating RNA of different lengths from the promoter start to t_{R1}. In one set, the deletions were nested with a common end at the *Hin*dIII site of pMZ245 (Figure 1). This set includes deletions A-1 and B-1 (Figure 1) and is defined as Δ SET 1 (represented by squares). In the other set, the deletions were nested with respect to the left end of deletion AA-2 (Figure 1). This set includes deletions AA-2 and BB-2 and is defined as Δ SET 2 (represented by circles). Galactokinase units were normalized to 100%, based on the galactokinase level at which t_{R1} was deleted. The nucleotide length was determined for each plasmid from the start of transcription to the t_{R1} terminator. In both sets, pMZ245 is represented with a length of 220 nucleotides (Figure 1).

Posttranscriptional Control of Gene Expression

The RNaseIII Operon: Suppressors of *era* Temperature-Sensitive Mutants. The *rnc* and *era* genes encode RNaseIII and Era, important functions involved in the life cycle of *E. coli*. RNaseIII is an enzyme that recognizes and cleaves double-stranded RNA structures. Mutants defective in RNaseIII have abnormal ribosome assembly and altered posttranscriptional expression. We have discussed the control of λ *N* gene translation by RNaseIII processing in previous Annual Reports. Within the RNaseIII operon, *era* is the second gene and is located just downstream of *rnc*. The *era* gene encodes a 35-kilodalton protein with a consensus G-domain similar to that of yeast RAS proteins. As expected, the purified protein has GTP-binding and GTPase activities.

We have isolated mutants of both RNaseIII and Era. Null mutants of RNaseIII are viable and have phenotypes similar to that of the classical *rnc105* mutant strain. Era is an essential function for *E. coli* viability, and a temperature-sensitive lethal Era mutant has been isolated, as described in previous Annual Reports. Other than its essential nature, the role of Era in the cell is not known. Initial studies to locate Era in the cell suggest that this protein is part of the cell membrane.

Suppressor mutants for the temperature-sensitive lethal *era* mutation have been isolated as temperature-resistant colonies following transposon mutagenesis with mini-Tn10-*kan*. Unlinked suppressor mutants often provide new information about how the suppressed gene is regulated or with what other genes it interacts. For this reason, we have initiated a study to determine the genetic location of the *kan* insertion suppressors and the identity of the affected gene(s). We have mapped two different suppressors, *ersB1* and *ersB2*, on the *E. coli* chromosome and have determined their nucleotide sequence. We performed genetic mapping with an ordered set of Tn10 tetracycline-resistant markers and found that both suppressors are P1 cotransducible, with *zgi203*::Tn10 and *zha6*::Tn10 located at about 69 minutes on the chromosome map. Fine-structure genetic, as well as restriction digest, mapping against the Kohara library of phages placed the mutations near the *rpoN* gene, which encodes a minor sigma factor (sigma-54) of RNA polymerase. RNA polymerase with sigma-54 specifically activates the transcription of genes involved in alternate nitrogen or carbon metabolism, and in some bacteria, it activates virulence and pathogenicity genes. Cloning and sequence analysis of the *kan* inserts and adjacent regions defined that *ersB2* is located within the *rpoN* gene itself and *ersB1* is within an open reading frame just two genes downstream of *rpoN* (Figure 3). We have determined the sequence of the entire 3-kilobase region between the *PstI* and *KpnI* sites on a plasmid containing these genes. This study demonstrates the precise location of the *ersB* insertions in *rpoN* and *orf4*. Beyond *orf4*, two "new" open reading frames are evident (Figure 3). Their reading frames overlap each other, as indicated in Figure 3, and they are transcribed in the same orientation as *rpoN-orf4*. In maxicell experiments using a plasmid carrying the entire segment shown in Figure 3, we have observed proteins of the size predicted for all the genes beyond and including *rpoN*.

The *ersB2* insertion creates a sigma-54 protein lacking its C-terminal 62 amino acids. Despite this absence, the sigma-54 mutant appears to carry out normal regulation of glutamine synthetase, whose gene is under sigma-54 transcriptional control. This observation suggests that the DNA-binding domain is not located in the C terminus of sigma-54 and supports current structure/function models. In other sigma factors, such as sigma-70, the DNA-binding domain is located in the C-terminal segment. If, as indicated, nitrogen regulation is normal in both *ersB* mutants, then the *ersB* mutations may affect *era* via mechanisms other than transcription regulation. This possibility is also supported by the observation that no DNA recognition site for sigma-54 is found near the promoter of *era* or within

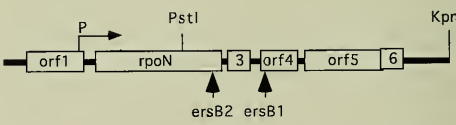


FIGURE 3. Structure of the *rpoN* region at 69 minutes in the *E. coli* chromosome. The DNA region (5 kilobases) shown is contained in a plasmid that complements bacterial suppressor mutants *ersB1* and *ersB2*. The sequence of the region is complete, with the genes shown as open rectangles. The positions of the *ersB*::*kan* insertions are indicated by vertical arrows. The location and direction of the *rpoN* promoter are indicated by P and a horizontal arrow.

the RNaseIII or neighboring *lep* operons. Complementation studies with plasmids carrying only the *rpoN* gene show that both *ersB1* and *ersB2* are complemented to restore temperature sensitivity. This complementation suggests that neither insertion is simply polar on a downstream gene. Experiments are in progress to define how these insertion mutations exert their effect on *era*.

Bacterial Protein Display Libraries

High-Resolution Epitope Mapping of HIV-1 RT. In collaboration with Stephen Hughes and Andrea Ferris (Gene Expression in Eukaryotes Section, Molecular Mechanisms of Carcinogenesis Laboratory), we have used a combination of two genetic studies to map the epitopes on HIV-1 RT recognized by monoclonal antibodies mAB28 and mAB42. One study mapped the epitopes by examining antibody recognition of the products of deletion derivatives of the genomic segment that encodes HIV-1 RT (Figure 4). The other study, described here, determined the sequence of peptides displayed on the surface of *E. coli* that were recognized by these two monoclonal antibodies. Both genetic approaches identified the same segment of RT. In the case of one of the two antibodies used in these analyses, mAB28, the structure of the antibody-RT complex was also examined by X-ray crystallography.

The binding of antibodies to their target proteins is one of the best-characterized protein-protein interactions. The nature of this interaction has been examined by a number of strategies; recently, phage peptide display libraries have been used to generate and identify the epitopes bound by an antibody. Instead of using phage for selecting peptides that bind to specific monoclonal antibodies, we have generated peptide display libraries on the surface of *E. coli*. These libraries differ in several ways from phage display libraries. Perhaps most importantly for this discussion, the random peptides displayed can be significantly longer than those accommodated by phage, and they are constrained within the sequence of the lambda receptor protein in the outer membrane of *E. coli* rather than appended to the end of a phage protein. The bacterial library of random peptides displayed on the surface of *E. coli* is similar to the one described in last year's Annual Report; this library consists of 6.8×10^6 independent clones containing peptide inserts in the lambda receptor that range in size from 9 to greater than 135 random amino acids.

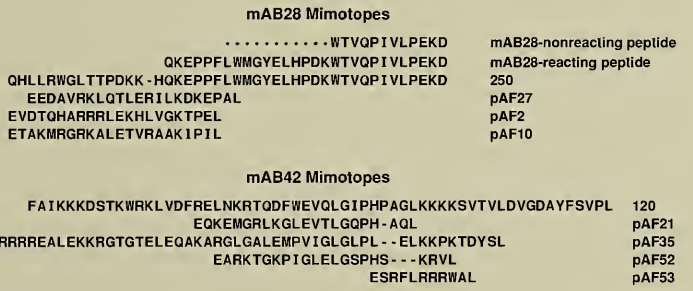


FIGURE 4. Mapping of HIV-1 RT epitopes. Epitopes for two monoclonal antibodies, mAB28 and mAB42, were analyzed. Below each antibody heading is the amino acid sequence of the region of RT containing the epitope that was defined using deletion mutagenesis; the number on the right indicates the position of the last amino acid. Two deletions of the N terminus are shown: one nonreacting and one reacting with mAB28. The pAF plasmids define synthetic mimotopes recognized by the monoclonal antibodies on the surface of *E. coli*. Computer programs were used to align the synthetic peptides with each other, and a consensus of this alignment was used to define the best match with the RT protein as shown in the figure. Dashes in the sequences indicate gaps. The filled circles in the mAB28-nonreacting sequence indicate that the same positions in the mAB28-reacting sequence are not likely to be part of the epitope as defined by the mimotopes below it. Presumably, only part of the epitope is present in the reacting peptide, i.e., that part that overlaps with the mimotopes.

We pursued the search for epitope analogs or mimotopes in two stages. The first stage, an enrichment procedure, relied on the ability of bacteria displaying a mimotope to adhere to paramagnetic beads coated with a monoclonal antibody. In the second stage, a colony screen was used to identify those peptides that bound the antibody most avidly.

We performed enrichments as described previously (1992 Annual Report), with the following modifications. Log-phase cultures of the bacterial populations were established, and expression of the recombinant surface proteins was induced for 15 minutes at 37°C. In the first cycle, magnetic beads coated with either of two types of monoclonal antibody (mAB28 or mAB42) were added at a ratio of approximately 1 bead per 10 bacteria and incubated 30 minutes at room temperature. Beads together with any adhering bacteria were recovered by magnetic separation, resuspended with YT broth, and spread on YT-ampicillin agar. Plates were incubated overnight at 37°C.

Subsequent cycles were performed in the following manner. The selected population of bacteria was grown on the agar surface and resuspended with M63 salts. Aliquots of the amplified bacterial suspension were frozen at –70°C with 15% glycerol for later analysis; one aliquot was regrown and expression of the recombinant surface proteins was induced as before. Bacteria from the first selection that adhere to structures other than the specific monoclonal antibody (either mAB28 or mAB42) were counterselected by incubating with beads coated with the other antibody (mAB42 or mAB28) for several minutes at room temperature at a ratio of approximately 1 bead per 10 bacteria. After the beads with adhering bacteria were removed by magnetic selection, the supernatant with nonadhering bacteria was incubated with beads coated with the original selecting monoclonal antibody for 15 minutes at room temperature at a ratio of approximately 1 bead per 25 bacteria. Beads with adhering bacteria were recovered by magnetic separation, resuspended with YT broth, and spread on YT-ampicillin agar. Cycles were repeated for further enrichment.

Distribution of the insert population from the enrichment cycles was monitored by polymerase chain reaction (PCR) analysis as described in last year's Annual Report. PCR amplification with primers that hybridized with vector sequences flanking the insert site generated a ladder of DNA sizes representing the number of oligonucleotides inserted. As members from the random population became enriched, the PCR product representing their size class of insert became more highly represented than the other size products and appeared as a denser band in the ladder. This analysis of the selected population was done following four cycles of enrichment with mAB28, and following five cycles with mAB42.

Individuals from populations following three, four, and five cycles of enrichment with mAB28, and following four and five cycles with mAB42, were screened for their ability to bind the respective monoclonal antibodies from diluted solutions. Nitrocellulose filters lifted from induced colonies from each population were exposed to the respective antibodies. To visualize positive colonies, we added goat anti-mouse IgG complexed with alkaline phosphatase. We screened 14 individuals originating from the product of five cycles of enrichment with mAB28 and identified 5 that bound the most dilute mAB28 preparation. Three isolates harbored the identical plasmid pAF2, and two harbored pAF10. From enrichment cycles three and four, pAF27, encoding the most avid mimotope, was recovered from ten independent colonies. Identical pAF27 isolates were recovered three times from the population representing four cycles of enrichment (a frequency of about 1/800); pAF27 was recovered six times from the population representing three cycles of enrichment (a frequency of about 1/7000). By comparing the peptide sequences of pAF27, pAF2, and pAF10, we assigned a consensus match that we then used to search the HIV-1 RT sequence. The best fit to the entire sequence is shown in Figure 4. The region of HIV-1 RT with homology to the consensus match coincides with the epitope location previously mapped for mAB28 by using deletion mutations.

Using dilute solutions of mAB42, we screened 1300 individuals originating from the product of five cycles of enrichment with mAB42, and 5400 originating from the product of four cycles with mAB42. We isolated four positive clones and determined their DNA sequence: pAF21 was identified in the product of five cycles of enrichment, and pAF35, pAF52, and pAF53 were identified in the product of four cycles of enrichment. For these four clones, we used the same computer analysis used for mAB28-derived plasmids. The best match to the entire protein is shown in Figure 4 .

Continuous Epitopes, Discontinuous Mimotopes? Both monoclonal antibodies (mAB28 and mAB42) recognized native and sodium dodecyl sulfate-denatured HIV-1 RT. The recombinant lambda receptors able to bind mAB42 were also able to bind as both denatured and native proteins. This is the expected result if a linear, contiguous epitope is displayed by the recombinant lambda receptors. Consistent with this result is the observation that the sequence homologous to RT appeared in different positions relative to the vector-encoded lambda receptor in the different recombinants. Unlike the recombinant lambda receptors able to bind mAB42, those that were able to bind mAB28 only bound as native proteins and the homologous sequence appeared in the same position relative to the vector-encoded protein. Thus, either the recombinant lambda receptors able to bind mAB28 must rely on the vector-encoded protein to maintain the insert in a functional conformation, or the vector is encoding a portion of the epitope. Since the flanking vector sequence is not similar to the epitope of HIV-1 RT, it is possible that another external portion of the lambda receptor is contributing to the epitope, generating a discontinuous mimotope.

Laboratory of Molecular Virology and Carcinogenesis

Stephen Oroszlan

Alan R. Rein
Nancy R. Rice

The Laboratory of Molecular Virology and Carcinogenesis is composed of the Molecular Biology of Retroviruses, Retroviral Genetics, and Immunochemistry Sections. Their common interest is in pathogenic retroviruses. Molecular studies are conducted on both oncogenic and nononcogenic retroviruses to understand their role in neoplasia, acquired immunodeficiency syndrome (AIDS), and other diseases. Toward this goal, they are studying the structure and expression of genes and gene products, the biosynthesis and biological function of viral structural and regulatory proteins, and enzymes as well as transforming proteins.

Under the direction of Stephen Oroszlan, the Immunochemistry Section has continued to focus on studies of the structure and function of retroviral protein components, including replication enzymes and regulatory proteins. The continued characterization of the retroviral protease has been of major interest together with studies on the antiviral effects of protease inhibitors.

The Retroviral Genetics Section, headed by Alan Rein, concentrates on genetic approaches to the analysis of retroviral biology and replication. In the current year, research in this section has focused on mechanisms of translational suppression, RNA packaging, and virus assembly.

The Molecular Biology of Retroviruses Section, headed by Nancy Rice, concentrates on studies of the *rel* oncogene, originally isolated as part of an acute leukemia retrovirus of young chickens. Current experiments have focused on interactions of the c-Rel protein with the members of the NF-κB family of transcription factors

Immunochemistry Section

Stephen Oroszlan

Carlton Briggs
Melonie B. Cassell
Deanna Gotte
Tadeusz Guszczynski
Catherine V. Hixson
Young D. Kim
Britta Köppe
Satoshi Kubota
Luis Menéndez-Arias
Vladimir Morozov
Karoly Nagy
Jozséf Tözsér
Tami Unangst-Steyskal
Kazunori Yoshimura
Matthew Young

The Immunochemistry Section has continued to concentrate on studies of the retroviral protease (PR), a virally encoded proteolytic enzyme that is essential for virus replication, as initially found by this laboratory for the PR of murine leukemia virus (MuLV).

The retroviral PR is a potential target for antiviral therapy. The PR of human immunodeficiency virus type 1 (HIV-1) has been studied most intensively, since HIV-1 is the causative agent of acquired immunodeficiency syndrome (AIDS), a fatal disease of epidemic proportions. In addition, we have investigated the PRs of other retroviruses that have been identified as potential human pathogens, e.g., human T-cell leukemia virus type I (HTLV-I), as well as retroviruses that are known to cause malignancies and other diseases in economically important domestic animals. We are interested in studying the biochemical and enzymological properties of HIV and other retroviral PRs to explore the possibility of designing broad-spectrum inhibitors as potential antiretroviral drugs. In previous years, we studied the substrate specificity of the PRs of HIV-1, HIV type 2 (HIV-2), bovine leukemia virus (BLV), mouse mammary tumor virus (MMTV), and equine infectious anemia virus (EIAV). This year, we report our progress in studies on the modeling and the substrate specificity of EIAV PR in comparison with that of HIV-1 PR, as well as the expression and characterization of the PR of Moloney-MuLV (Mo-MuLV).

One of the functions of the retroviral PR is the processing of the Gag and Gag-Pol polyproteins into smaller functional units during the late phase of virus replication, whereby the noninfectious immature particles are transformed into mature infectious viruses. Previously reported results suggested that the PR also functions in the early phase of the virus life cycle, as we had originally proposed, based on in vitro studies with purified capsids of EIAV. In this report, we describe studies on the antiviral effect of PR inhibitors in a single cycle of infection by HIV-1. We have obtained evidence that the nucleocapsid (NC) protein is cleaved in EIAV capsids during endogenous reverse transcription. This cleavage is blocked by PR inhibitors and may be required for the function and/or transport of the preintegration (cDNA-protein) complex.

We also report studies using affinity chromatography and amino acid microsequencing techniques that led to the definitive demonstration that the Rex protein, the posttranslational regulator of HTLV-1, specifically binds to the nucleolar shuttle protein B-23. This binding suggests that B-23 functions as a shuttle not only for the nucleolar transport of ribosomal components but also for the transport of retroviral mRNAs containing the Rex-responsive element (RXRE) from the nucleus to the cytoplasm and the Rex protein from the cytoplasm to the nucleolus.

In the last part of this report, we describe the expression and characterization of the MMTV transframe protein (p30), which contains the nucleocapsid protein derived from *gag* as its N-terminal domain and the 5′ *pro*-encoded sequence as its C-terminal domain. Furthermore, we demonstrate that this fusion protein has dUTPase activity.

Molecular Model of the EIAV PR

We built a three-dimensional model of EIAV PR based on the determined crystal structures of the related Rous sarcoma virus (RSV) and HIV-1 PRs, in collaboration with Irene T. Weber (Jefferson Cancer Institute, Philadelphia, PA). This model has helped us in interpreting the results of our kinetic studies performed to characterize the specificity of EIAV PR. Based on the refined crystal structure of HIV-1 PR with inhibitor JG365, the molecular modeling of EIAV PR was performed using FRODO on an Evans and Sutherland PS390 and subsequently on an ESV10. The FRODO idealization routine REFI was run to maintain reasonable geometry. The region leading to and including the flaps was modeled by analogy to RSV PR.

The 104-residue EIAV PR has 30 identical and 11 similar amino acids compared to HIV-1 PR and 25 identical and 18 similar amino acids compared to RSV PR (Figure 1). The overall structure of EIAV PR is predicted to be close to that of HIV-1 PR (Figure 2). Two regions show significant differences, however: there are 6 additional residues leading to the tip of the flap, which is predicted to be involved in interactions with the substrate, and there is a single-residue deletion in the β b′ strand at a position equivalent to residue 60 in HIV-1 PR (Figure 1). The conformation of the residues leading to the flap was modeled by analogy to the corresponding region of RSV PR. The peptide substrate VSQNYPIVQ was modeled by analogy to the inhibitors in the cocrystal structures of EIAV and HIV-1 PRs, and the residues forming the substrate-binding sites of EIAV PR were identified. Only 5 residues forming the subsites for predicted substrate binding differ between the two enzymes (Table 1).

FIGURE 1. Sequence comparison of HIV-1, HIV-2, EIAV, and RSV PRs based on structural alignment. The elements of secondary structure observed for the crystal structures of HIV-1 and RSV PRs are indicated by letters a to d and q for β-strands and by h for helix. The dashed line indicates the residues in the flap that were not visible in the crystal structure of RSV PR.

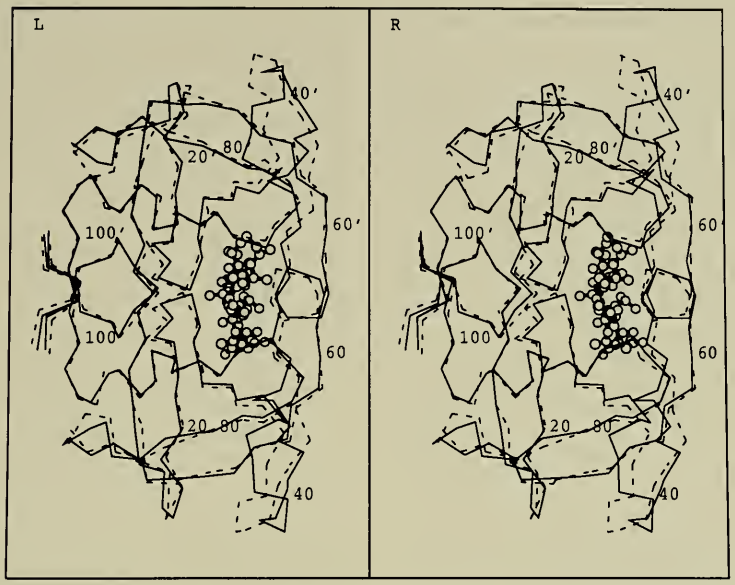


FIGURE 2. Stereoview of the EIAV PR structure, showing superposition of C_α atoms in the model structure of the EIAV PR dimer (solid line) with the HIV-1 PR crystal structure (dashed line). The inhibitor from the cocrystal structure with HIV-1 PR is shown in a ball-and-stick representation. Modeled according to the RSV PR, the region of EIAV PR leading to the flap differs from that of HIV-1 PR.

TABLE I. Residues forming the subsites of EIAV and HIV-1 PRs

Subsite	HIV-1/EIAV residues
S4	Arg8′, Asp29, **Asp/Thr30**, (**Gly50, Thr51, Gly52**), Ile47/53, **Thr74/Arg79**, Leu76/81
S3	Arg8′, Leu23′, Asp29, **Gly48/Ile54**, Gly50/55, Pro81′/86′, Val82′/87′
S2	Ala28, Asp29, **Asp/Thr30**, Val32, Ile47/53, **Gly48/Ile54**, Gly49/55, **Ile50′/Val56′**, Leu76/81, Ile84/89
S1	Arg8′, Leu23′, Asp25′, Asp25, Gly27, **Gly48/Ile54**, Gly49/55, **Ile50/Val56**, **Thr80′/Ile85′**, Pro81′/86′, Val82′/87′, Ile84′/89′
S1′	Arg8, Leu23, Asp25, Asp25′, Gly27′, **Gly48′/Ile54′**, Gly49′/55′, **Ile50′/Val56′**, **Thr80/Ile85**, Pro81/86, Val82/87, Ile84/89
S2′	Ala28′, Asp29′, **Asp/Thr30′**, Val32′, Ile47′/53′, **Gly48′/Ile54′**, Gly49′/55′, **Ile50/Val56**, Leu76′/81′, Ile84′/89′
S3′	Arg8, Leu23, Asp29′, **Gly48′/Ile54′**, Gly50′/55′, Pro81/86, Val82/87
S4′	Arg8, Asp29′, **Asp/Thr30′**, (**Gly50′, Thr51′, Gly52′**), Ile47′/53′, **Thr74/Arg79′**, Leu76′/81′

Amino acid residues in the second subunit of the dimer are indicated by a prime. The residues that differ in the two PRs are indicated in boldface as HIV-1 PR residue/EIAV PR residue. The EIAV PR residue numbers change after residue 36 due to an insertion relative to HIV-1 PR. The residues in the extra loop of the EIAV PR flap are in parentheses.

Comparison of the Substrate Specificity of EIAV and HIV PRs

Last year, we reported the purification and initial characterization of the substrate specificity of EIAV PR and the comparison of its specificity to that of HIV-1 PR. We have continued these studies using a complete series of oligopeptide substrates with single amino acid substitutions (Table 2). The PR assays were performed by the high-pressure liquid chromatography (HPLC) point assay. Relative activities were calculated from the molar amounts of peptides cleaved per unit time by dividing the activity on a given peptide by the activity on the unmodified substrate VSQNYPIVQ, at less than 20% substrate turnover. Based on the average values calculated from measurements performed in duplicate, the error was less than 10%. As expected, the activities of EIAV and HIV-1 PRs were similar for substitutions at most positions (Table 2). Substantial differences are observed in the relative activities of the two enzymes for peptides with substitutions at the P4 and P2 positions.

At the S4 binding pocket, the substitution of Asp30 to Thr as well as the presence of Arg79 actually turns the negatively charged pocket into a positively charged one (Table 1). When the relative activities of EIAV and HIV-1 PRs are compared for substitutions of the P4 position, it is clear that the hydrophobic residues Leu, Val, Phe, Ile, and Ala form better substrates for EIAV PR than for HIV-1 PR (Table 2). This effect is partly due to the additional residues 49 to 52 in the flaps of EIAV PR, which make subsite S4 more enclosed (Figure 3). EIAV PR also has the more hydrophobic Thr instead of Asp30 in HIV PR, and hydrophobic residues at P4 can interact with the aliphatic side chain of Arg79. Gly at P4, which lacks a side chain, is a poor substrate for EIAV PR, whereas Gly P4 makes a good substrate for the less-enclosed S4 subsite of HIV-1 PR.

Subsite S2 is illustrated in Figure 4. Subsites S2 and S2′ show the nonconservative change of Asp30 to Thr in EIAV PR (Table 1), which makes the site more hydrophobic and less negatively charged. Substitution of Ile50 to Val56 introduces a smaller hydrophobic side chain, and substitution of Gly48 to Ile54 introduces a larger hydrophobic side chain that is probably directed away from the subsite (Figure 4). These changes suggest that somewhat larger and more hydrophobic side chains may

TABLE 2. Relative activities obtained with EIAV and HIV-1* PRs for substrates having single amino acid substitutions from positions P4 to P3′†

Substitutions	Ser P4		Gln P3		Asn P2		Tyr P1		Pro P1′		Ile P2′		Val P3′	
Ala	1.2	(0.3)	0.14	(0.34)	5.1	(0.53)	0	(<0.01)	0	(0)‡	0.06	(0.17)‡	0.24	(0.34)
Leu	1.5	(0.02)	<0.01	(0.40)	0.74	(0.06)	0.09	(0.21)	<0.01	(0)	0.24	(0.44)‡	5.2	(3.2)
Val	1.4	(0.05)	<0.01	(0.71)	2.5	(0.17)	0	(0)	<0.01	(0)	1.0	(0.93)‡		
Ile	1.2	(0.03)			1.3	(0.1)	0	(0)						
Phe	1.3	(0.03)	0.23	(0.22)	0.16	(0.03)	1.4	(1.7)	<0.01	(0)	0.06	(0.07)‡	5.9	(2.4)
Gly	0.12	(0.76)	0.21	(0.18)	0.09	(0.12)	0	(0)			<0.01	(<0.01)‡	0.10	(0.24)
Ser							0	(0)	0	(0)			0.27	(0.36)
Thr					3.24	(0.02)							0.10	(0.40)
Met	0.65	(0.04)					0.15	(0.23)	<0.01	(0)				
Asp			<0.01	(<0.01)	0.15	(0.13)	0	(0)			<0.01	(<0.01)‡	<0.01	(0.05)
Glu													0.16	(0.1)
Gln											<0.01	(0.03)	0.73	(0.67)
Lys	0.25	(0.06)	0.08	(0.12)	0	(0)	0	(0)	0	(0)‡	0	(0)‡	0.09	(0.1)
Arg	0.14	(0.15)											0.41	(0.57)
Cys					7.3	(0.92)	0.09	(<0.01)						

*Numbers in parentheses.

†Relative activities were calculated by determining the molar amount of peptide cleaved and dividing the activity on a given peptide by the activity obtained with the unmodified substrate Val-Ser-Gln-Asn-Tyr*Pro-Ile-Val-Gln-NH_2 (SP-211) at 0.4 mM concentration.

‡Taken from Bláha et al. (Int J Pept Prot Res 38:453, 1991).

be accommodated at P2 and P2′ in EIAV PR compared to HIV PR. In fact, at P2, the medium to small hydrophobic residues Ala, Val, Ile, Cys, and Leu are substantially better substrates of EIAV PR than of HIV-I PR, as is Thr, which is partly hydrophobic. The β-branched side chains of Val, Ile, and Thr are accommodated better in the S2 and S2′ subsites of EIAV PR due to the presence of the smaller Val56 instead of the larger Ile50 of HIV-1 PR. Phe is too large to be a good substrate at P2 or P2′ for either enzyme. At P2′, Val and Ile form the best substrates for both enzymes, while the substitution of Ala or Leu results in poorer substrates for EIAV PR compared to HIV PR.

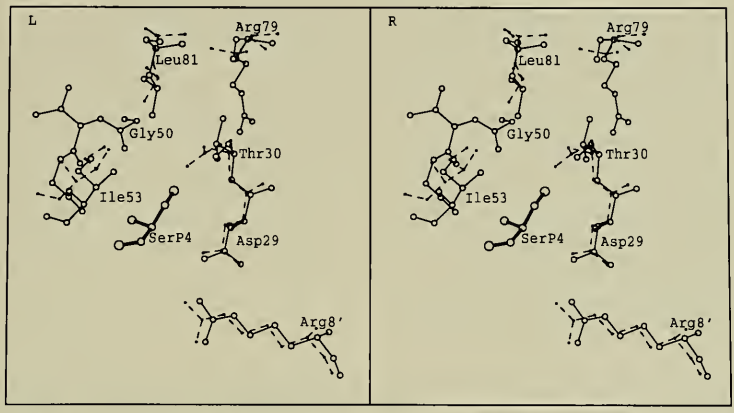


FIGURE 3. Stereoview of the S4 subsite of EIAV and HIV-1 PRs. The S4 subsite of EIAV PR (solid line) and HIV-1 PR (dashed line) is shown with Ser at P4 of the peptide substrate.

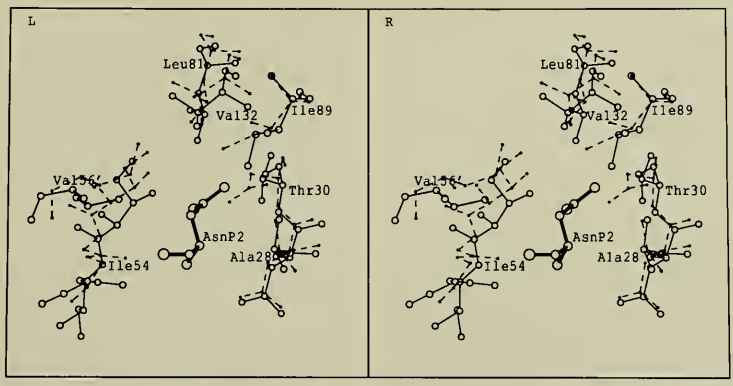


FIGURE 4. Stereoview of the S2 subsite of EIAV and HIV-1 PRs. The S2 subsite of EIAV PR (solid line) and HIV-1 PR (dashed line) is shown with Asn at P2 of the peptide substrate.

Hydrolysis of Oligopeptides Having the Sequence of NC Protein Cleavage Sites

Oligopeptides having the sequence of the cleavage site in the first and second cysteine arrays of EIAV NC protein (EIAV NC-1 and NC-2) were studied as substrates for EIAV and HIV PRs (Table 3). These peptides were hydrolyzed at the expected -Cys*Tyr- and -Cys*Phe- bonds. SP-346, a specific retroviral PR inhibitor, inhibited these cleavages. The kinetic parameters for the hydrolysis of these peptides are listed in Table 3. A longer peptide representing the C-terminal part of the NC protein, including the complete second Cys array (APKVC*FKCKQPGHFSKQCRSVPKNGKQGAQGR-PQKQTF), was also synthesized and found to be hydrolyzed by EIAV PR at the -Cys*Phe- bond (data not shown). On the other hand, none of the EIAV "NC peptides" was cleaved by either HIV-1 or HIV-2 PRs (k_{cat} values are estimated to be less than 0.01 s^{-1}). This result is predicted to be due at least partly to the large difference for the preference at the P4 and P2 positions: both Thr and Val are far better accommodated in the S2 pocket of EIAV PR than in that of HIV-1 PR, while hydrophobic residues such as Ala, Val, and Ile are preferred in S4 (Table 2). Concerning the S4 subsite, results with HIV-2 PR were very similar to those found with HIV-1 PR (not shown).

Oligopeptides representing analogous regions in the NC proteins of the HIVs (HIV-1 NC-1 and HIV-2 NC-1, Table 3) were also found to be hydrolyzed by HIV and EIAV PRs; however, cleavage occurred at the -Phe*Asn- and -Trp*Asn- bonds instead of the predicted (see the 1990 Annual Report) -Cys*Phe- and -Cys*Trp- bonds. This "shift" may be partially due to the presence of Lys adjacent to Cys. Substitution of Asn with Lys at the P2 position of substrate SP-211 prevented hydrolysis by both HIV and EIAV PRs (Table 2). On the other hand, while replacement of the P1 Tyr to Cys produced an inefficient substrate for both HIV-1 and EIAV PRs at the P2 position, Cys fits very well in the pocket of EIAV and HIV PRs and formed a better substrate for EIAV PR compared to HIV-1 PR. However, it should be pointed out that intramolecular interactions between adjacent side chains of the substrates may also contribute to the specificity at a given subsite, as found for HIV-1 PR.

TABLE 3. Proteolytic processing of oligopeptides representing determined or predicted cleavage sites in the NC proteins of EIAV, HIV-1, and HIV-2

		EIAV-PR	
Location of cleavage site	Sequence of peptide	K_m (mM)	k_{cat} (s^{-1})
EIAV NC-1*	AAQTC*YNCGK	0.35	0.2†
EIAV NC-2‡	APKVC*FKCKQ	0.6	0.2†
HIV-1 NC-1*	KMVKCF*NCGK	0.47	1.2§
HIV-2 NC-1*	KPIKCW*NCGK	0.02	0.02¶

*Cleavage site in the first Cys array (Roberts et al., Prot Eng 4: 695, 1991).

†These peptides were not hydrolyzed by the HIV-1 PR or HIV-2 PR.

‡Cleavage site in the second Cys array (Roberts et al., Prot Eng 4: 695, 1991).

§This peptide was also hydrolyzed by HIV-1 PR, with a k_{cat} value of 0.01.

¶This peptide was also hydrolyzed by HIV-2 PR, with a k_{cat} value of 0.01.

As an overall conclusion, our previous and new results suggest that lentiviral PRs share common characteristics in their specificities. HIV-1 and HIV-2 PRs had shown similar specificities on various oligopeptides in our assay system (see the 1990 and 1991 Annual Reports). As the divergence is increased between two PRs, one would expect more substantial differences in the kinetics of substrate hydrolysis, or even shifts in the site of cleavage. The PR of EIAV, a lentivirus, is still a close relative of the HIV PRs, and only 5 of 20 residues differ in the predicted substrate-binding region. In fact, these enzymes share many characteristics in specificity. Although some of the peptides that we studied were hydrolyzed only by EIAV PR (suggesting its broader specificity), if a peptide was hydrolyzed by both enzymes, the cleavage took place at the same bond. The molecular explanation of this phenomenon lies in the nature of the enzyme-substrate interaction. Since about seven consecutive substrate residues are recognized by the enzyme, changes in one or a few binding subsites of the enzyme may alter, but do not necessarily abolish, the enzyme activity. This effect should be very advantageous for an enzyme encoded by rapidly mutating RNA.

Bacterial Expression and Characterization of Mo-MuLV PR

Last year, we reported the purification and characterization of MMTV PR expressed in *Escherichia coli*, using pGEX-2T as an expression vector. This year, we have expressed and purified Mo-MuLV PR and compared this PR with MMTV PR in terms of substrate specificity and enzymological properties.

The formation of infectious Mo-MuLV involves the proteolytic processing of Gag ($Pr65^{gag}$), Gag-Pol ($Pr180^{gag\text{-}pol}$), and Env ($gPr85^{env}$). As we initially determined, in Mo-MuLV the virally encoded PR appears to be responsible for cleavages not only in the Gag-related precursors, but also in Env, where cleavage of the cytoplasmic domain of Mo-MuLV transmembrane protein (TM) has been attributed to the retroviral PR.

Despite early work leading to the isolation of Mo-MuLV PR from the virus, studies on the enzymatic properties of this enzyme have been hampered by the low yields obtained and the lack of alternative sources (e.g., PR expressed in bacteria).

We amplified the Mo-MuLV PR gene by the polymerase chain reaction (PCR), using as a template a plasmid containing part of the *gag-pol* gene from an infectious clone. One of the synthetic primers we used was 5′-GCGCGGATCCACCCTAGATGACCAGGGAGGTCAG-3′, which corresponds to the 5′ end of the PR gene. The underlined glutamine codon was inserted in place of TAG, since PR synthesis in Mo-MuLV arises from the in-frame suppression of the *gag* terminator. The obtained DNA was cloned into pGEX-2T, a plasmid containing the glutathione S-transferase (GST) gene of *Schistosoma japonicum*, and a nucleotide sequence encoding a thrombin cleavage site. We found that the nucleotide sequence of the insert in a selected clone (pMuLVPR3.2) was identical to that of the template, except for the introduced mutation of the *gag* terminator.

We used the clone construct pMuLVPR3.2 for expression and purification of Mo-MuLV PR. After induction with isopropyl-β-thiogalactopyranoside (IPTG), cells expressing the chimeric protein GST-MuLV PR were harvested, lysed, and centrifuged. We purified the PR after thrombin cleavage of the chimeric protein bound to a glutathione-Sepharose 4B support. Polyacrylamide gel electrophoresis in the presence of sodium dodecyl sulfate (SDS-PAGE) and Western blot analysis of the purified enzyme are shown in Figure 5 (lanes 2 and 5, respectively). The yield after the purification procedure was approximately 0.2 mg/liter of culture (2.5 g of *E. coli* cells, wet weight). The specific activity of the purified viral enzyme was 2100 pmol/min•μg of enzyme, using the peptide TSQAFPLRAG as a substrate.

The optimum pH of the Mo-MuLV PR activity was 6.0, and increasing the ionic strength (up to 3 M) seemed to have a stimulatory effect on this activity. We tested a series of aspartyl PR-specific inhibitors for their ability to inhibit Mo-MuLV PR (Table 4).

Synthetic oligopeptides mimicking the maturation sites found in the Gag polyprotein of Mo-MuLV were synthesized and used as substrates in proteolytic assays using the recombinant Mo-MuLV PR. Kinetic parameters calculated using these substrates are shown in Table 5. In addition to maturation sites in Gag-related polyproteins, cleavage of the cytoplasmic domain of the TM protein has been attributed to the viral protease

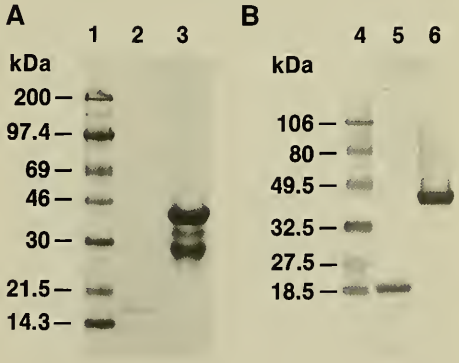


FIGURE 5. SDS-PAGE and Western blot analysis of Mo-MuLV PR, purified and in its chimeric form. (A) Coomassie blue-stained SDS-polyacrylamide gel of the purified recombinant Mo-MuLV PR (lane 2) and the precursor fusion protein, obtained after elution from a glutathione-agarose column (lane 3). Protein molecular mass markers are shown in lane 1 for comparison. (B) Western blot of the purified PR (lane 5) and the chimeric precursor (lane 6) compared to prestained markers (lane 4).

TABLE 4. Effect of aspartyl-PR inhibitors on the in vitro activity of Mo-MuLV PR

Inhibitor	Reference	IC_{50} (μM)
KH164	Hui et al., FASEB J 5:2606, 1991	0.062 (2.1)*
	Lai et al., J Acquir Immune Defic Syndr 6:24, 1993	
Compound 3	Grobelny et al., Biochem Biophys Res Commun 169:1111, 1990	0.134 (0.09)
UK-88,947	Baboonian et al., Biochem Biophys Res Commun 179:17, 1991	2.2 (4.3)
Compound 4	Grobelny et al., Biochem Biophys Res Commun 169:1111, 1990	6.7
Acetyl-pepstatin A	Richards et al., FEBS Lett 247:113, 1989	25.0†

*The IC_{50} values for these compounds were also determined in the assay conditions previously established for MMTV PR (Menéndez-Arias et al., J Biol Chem 267:24134, 1992), using the oligopeptide LTFTFPVVFMRR (SP-1287) as a substrate. The values obtained under these conditions are shown in parentheses.

†The IC_{50} was 10 μM when assayed in 0.25 M potassium phosphate buffer (pH 5.0) containing 3 M NaCl, using the oligopeptide TSQAFPLRAG (SP-2209) as a substrate.

(Crawford and Goff, J Virol 53:899, 1985; Katoh et al., Virology 145:280, 1985; Schultz and Rein, Virology 145:335, 1985). We synthesized peptides mimicking such a cleavage site found between p15E and p2E. After subjecting these peptides to proteolytic assays, we found that two of them, VQAL*VLTQ (SP-1435) and VQAL*VLTQQRR (SP-1438), were cleaved at the expected position and showed K_m and k_{cat} values similar to those observed for other peptides representing maturation sites in Gag and Gag-Pol (Table 5). Furthermore, complete inhibition of these cleavages was achieved by including in the assay the inhibitor KH164 (see Table 4) at a final concentration of 1 to 10 μM.

Cleavage by Mo-MuLV PR of substrates corresponding to maturation sites found in precursor polyproteins of other retroviruses revealed its broad specificity, a property that is shared by other retroviral PRs. Interestingly, we found that peptides mimicking the cleavage site at the N terminus of the capsid portion of both HIV-1 and Mo-MuLV Gag proteins were poor substrates of the MMTV PR, but were rapidly hydrolyzed by Mo-MuLV PR. These data suggest that both Mo-MuLV and HIV-1 PRs are similar in terms of substrate specificity, while MMTV PR shows significant differences in the substrate-binding sites.

This hypothesis is also supported by data obtained using aspartyl-PR inhibitors. We found that the IC_{50} value for MMTV PR was 1.2 μM when compound 3 was used as an inhibitor, while KH164 and UK-88,947 were weak inhibitors of this enzyme (data not shown). In contrast, the IC_{50} values obtained for Mo-MuLV PR when tested in the same conditions as MMTV PR were at least ten times lower, suggesting that HIV-1 PR inhibitors are more effective towards Mo-MuLV PR than towards MMTV PR.

The analysis of the substrate specificity of retroviral PRs has potential applications in the design of new broad-spectrum antiretroviral drugs. Furthermore, the relationship found between HIV-1 and Mo-MuLV PRs provides a good model for studying the enzyme-substrate interactions of these enzymes. We are continuing a kinetic analysis of Mo-MuLV PR, using the oligopeptide substrate SP-211 (VSQNY*PIVQ) and a series of analogs containing substitutions in positions P_4-P_3'.

The Role of HIV-1 and Other Retroviral PRs in the Early Phase of Virus Replication

We suggested the possible involvement of retroviral PR in the early stages of the viral replication cycle on the basis of in vitro experiments with purified capsids of EIAV. These studies demonstrated a novel protein processing pathway that is mediated by the viral PR and involves the regulated in situ cleavage of the NC protein assembled with the viral RNA. It was proposed that such a proteolytic event may be required for virus replication. Furthermore, we have observed that a specific inhibitor of HIV-1 PR, UK-88,947, not only inhibited HIV-1 maturation in the late phase but apparently also prevented or interfered with cDNA synthesis if added to

TABLE 5. Kinetic parameters of the purified recombinant Mo-MuLV PR with different synthetic peptide substrates

Peptide	Virus	Polyprotein cleavage site*	Peptide sequence†	K_m (μM)	$k_{cat} \times 10^3$ (s^{-1})	k_{cat}/K_m ($M^{-1}s^{-1}$)
SP-76	Mo-MuLV	MA/p12	PRSSLY*PALTP	80.4 ± 11.6	163.8 ± 6.6	2037.3
SP-2209	Mo-MuLV	p12/CA	TSQAF*PLRAG	98.0 ± 15.1	714.7 ± 32.5	7292.9
SP-2210	Mo-MuLV	CA/NC	MSKLL*ATVVS	29.9 ± 12.1‡	439.2 ± 50.0	14689.0
SP-2211	Mo-MuLV	NC/PR	TQTSLL*TLDDQ	44.5 ± 12.2	114.4 ± 12.9	2570.8
SP-617	Mo-MuLV	PR/RT	PLQVL*TLNIERR	16.6 ± 8.2	70.0 ± 8.1	4216.9
SP-2213	Mo-MuLV	RT/IN	TSTLL*IENSS	68.0 ± 16.6	294.2 ± 19.8	4326.5
SP-1435	Mo-MuLV	p15E/p2E	VQAL*VLTQ§	24.2 ± 10.9‡	279.0 ± 32.4	11528.9
SP-520	MPMV¶	p12/CA	PKDIF*PVTET	510.4 ± 88.8	1275.4 ± 47.5	2498.8
SP-2231	BLV	MA/CA	PPAIL*PIISE	39.1 ± 18.1‡	114.1 ± 24.8	2918.2
SP-2065	EIAV	MA/CA	PSEEY*PIMID	120.6 ± 22.9	415.2 ± 142.8	3442.8
SP-2079	HTLV-I	MA/CA	APQVL*PVMHP	150.1 ± 18.5	96.4 ± 11.9	642.2
SP-211	HIV-1	MA/CA	VSQNY*PIVQ	260.9 ± 25.5	551.2 ± 16.7	2112.7
SP-1287	MMTV	n/CA	LTFTF*PVVFMRR	19.2 ± 3.7	88.1 ± 7.1	4588.5
SP-540	MMTV	MA/pp21	SDLVL*LSAEARR	43.5 ± 5.3	61.1 ± 2.1	1404.6
SP-2201	MMTV	pp21/p3	DSKAF*LATDW	200.1 ± 68.3	237.8 ± 51.1	1188.4
SP-2202	MMTV	p3/p8	DELIL*PVKRK	84.6 ± 16.1	209.6 ± 15.2	2477.5
SP-1342	MMTV	p8/n	VGFAG*AMAEAR	166.3 ± 72.9‡	8.0 ± 2.4	48.1

*Viral proteins are designated according to the nomenclature of Leis et al., J Virol 62:1808, 1988: MA, matrix protein; CA, capsid protein; NC, nucleocapsid protein; PR, protease; RT, reverse transcriptase; IN, integrase.

†All peptides have been synthesized with a free N terminus and an amide at the C terminus, except those representing MMTV and MPMV substrates, which have both N and C termini free. Large asterisks indicate the observed cleavage after incubation with the recombinant protease. Underlined arginines were added to increase the solubility of the substrate and are not present in the corresponding viral sequences.

‡An inhibitory effect was observed when the substrate concentration was well above the apparent K_m value.

§Assays using this substrate were performed in the presence of 1% DMSO.

¶Mason-Pfizer monkey virus.

the cells prior to infection. In contrast, it was reported by others (Jacobsen et al., J Virol 66:5087, 1992) that another potent inhibitor of HIV-1 PR, Ro 31-8959, did not restrict progression of the early steps of HIV-1 replication. However, these seemingly contradictory observations were made either by using amplification of cDNA by semiquantitative PCR with a single primer pair (5′ *gag*) at a single time point (18 hours) after infection or by using only a single dose (1 μM) of PR inhibitor.

We are conducting experiments designed to study primary HIV-1 infection of different cell lines in the presence and absence of the inhibitor UK-88,947, as well as Ro 31-8959, which was resynthesized (in the laboratories of Pfizer Central Research) according to published methods. We used PCR amplification with a panel of oligonucleotide primer pairs recognizing various regions of the viral genome. In addition, we carried out quantitative infectivity assays in a single cycle of infection with the HIV-1_{IIIB} wild-type strain and a replication-defective virus, HIV-*gpt*(HXB-2), in the presence and absence of the same inhibitors. Our results demonstrate a definitive dose- and time-dependent antiviral effect of these inhibitors under conditions that do not allow secondary infection by progeny virus.

PCR Amplification of HIV-1 cDNA with Various Primers

To characterize the effect of PR inhibitors on the progression of HIV-1 cDNA synthesis, lysates of H9 cells treated with UK-88,947 1 hour prior to infection with

HIV-1_{IIIB} were analyzed at 4 hours and 18 hours postinfection for cDNA by PCR amplification using a panel of primer sets (Figure 6). In addition to the *gag*-specific primers (JA4, JA7), primer sets specific to R/U5 (M667 and AA55), *nef*/U3 (AA943 and AA946), and long terminal repeat (LTR)/*gag* (M667 and M661) were used. At 4 hours postinfection, cDNA was detected with all four primer pairs, indicating that the virus had entered the cell and reverse transcription has been initiated in the presence of the PR inhibitor. Surprisingly, after 18 hours, only the R/U5- and the *nef*/U3-specific primers were able to detect cDNA. The *gag* and LTR/*gag*-specific primers were unable to amplify cDNA. These results suggest that the PR inhibitor did not block initiation and progression of reverse transcription. However, by 18 hours postinfection, the full-size cDNA detected at 4 hours in the inhibitor-treated cells may have been degraded instead of being integrated in the presence of UK-88,947.

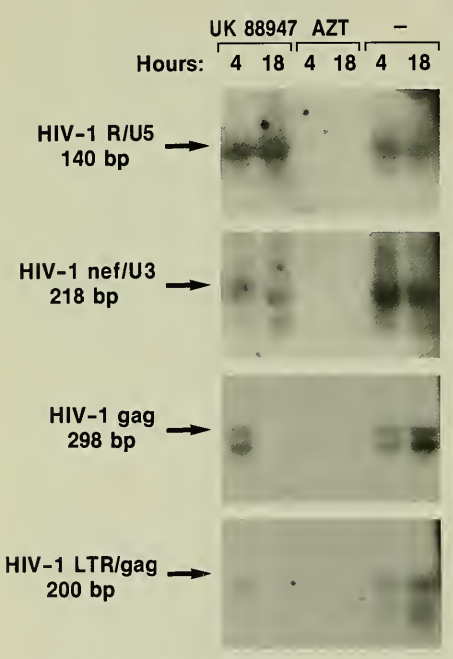


FIGURE 6. Hybridization analysis of HIV-1 proviral DNA synthesis in acutely infected H9 cells in the presence of PR inhibitor UK-88,947. H9 cells were treated with UK-88,947 (24 μM) or AZT (8 μM) 1 hour prior to infection with HIV-1_{IIIB} (0.1 moi). Cells were harvested at 4 and 18 hours postinfection, then DNA (from 10^5 cells) was subjected to 35 cycles of PCR amplification with R/U5, *nef*/U3-, *gag*-, or LTR/*gag*-specific primers. The minus symbol indicates untreated, infected H9 cells. Samples at 0 hour were not analyzed by PCR and Southern hybridization. In the AZT-treated samples, very little (see R/U5, 4-hour lane) or no cDNA was detected.

Antiviral Effect of PR Inhibitors in a Single Cycle of Infection by HIV-1_{IIIB} and HIV-*gpt*(HXB-2)

We used two assay systems to quantitate HIV-1 infectivity in a single cycle of infection in the presence and absence of PR inhibitor. First, the effect of UK-88,947 and Ro 31-8959 on primary infection was quantitatively determined in the genetic assay system (Page et al., J Virol 64:5270, 1990), which uses the replication-defective virus HIV-*gpt*(HXB-2). Infection of HeLaT4 cells leads to the formation of drug-resistant colonies after 10 to 12 days of growth in the selection medium. The number of HeLaT4 colonies is proportional to the titer of input virus and secondary infection is prevented because the progeny virus is not infectious (Page et al., J. Virol. 64: 5270, 1990). Treatment of HeLaT4 cells with UK-88947 or Ro 31-8959 1 hour prior to infection with HIV-*gpt* (HXB-2) resulted in the reduction of the number of drug-resistant (infected) colonies in a dose-dependent manner (Table 6).

When added to the cultures 1 hour prior to the addition of virus, UK-88,947 inhibited the formation of infected HeLaT4 colonies with an IC_{50} of 3.6 μM and IC_{90} of ~24 μM. For Ro 31-8959, the IC_{50} was 4.6 μM and IC_{90} was 18 μM. However, the additions of the PR inhibitors to the HeLaT4 cultures 2.5 hours after infection did not prevent formation of colonies (i.e., infection).

We also studied HIV infectivity in the presence and absence of PR inhibitors in the MAGI assay (Kimpton and Emerman, J Virol 66:2232, 1992), which is based on the activation of the integrated β-galactosidase gene of HeLaCD4-LTR/β-gal cells by HIV Tat protein. UK-88,947 and Ro 31-8959 effectively inhibited infection with both HIV-*gpt*(HXB-2) and HIV-1_{IIIB} when added to the cells 1 hour prior to infection. UK-88,947 had an

TABLE 6. Effect of PR inhibitors UK 88,947 and Ro 31-8959 on a single cycle of infection of HeLaT4 cells with HIV-*gpt* (HXB-2)*

Inhibitor concentration† (μM)	Percent inhibition	IC_{50}	IC_{90}
UK-88,947			
6	59.3	3.6	~24
12	76.3		
24	85.8		
Ro 31-8959			
1	31	4.4	18.05
5	55		
10	82.5		
20	91.8		

*HIV-*gpt*(HXB-2) was produced by cotransfection of COS-7 cells with HIV-*gpt* plasmid DNA and HXB-2 *env* DNA (1.2×10^4 IU/ml). HeLaT4 cells (2×10^5) were infected with the virus (0.1 moi) for 1 hour.

†Inhibitor was added to the cultures 1 hour prior to infection with virus. At 18 hours postinfection, selection medium containing mycophenolic acid (50 μg/ml) was added to the cultures. Colonies of drug-resistant cells were fixed, stained, and counted at 12 days postinfection. When added at −1 hour, AZT (8 μM) inhibited colony formation by 95%. IC_{50} and IC_{90} concentrations were determined by Hill equation analysis (confidentiality limit, 95%; goodness of fit index, 0.35). Stock solutions of inhibitors were made in DMSO. The highest concentration of DMSO in the culture fluid was 0.2%.

IC_{50} of 15.8 μM and 12.6 μM and an IC_{90} of 25.5 μM and 24.6 μM, respectively, against the two viruses, while Ro 31-8959 had an IC_{50} of 3.1 μM and 4.3 μM and an IC_{90} of 12.8 μM and 10.2 μM, respectively. Addition of the compounds 15 minutes after infection also inhibited replication of HIV-1_{IIIB}, whereas at 30 minutes after infection the addition had minimal inhibitory effect (Table 7).

TABLE 7. Effect of PR inhibitors UK-88,947 and Ro 31-8959 on a single cycle of infection of HeLaCD4-LTR/β-gal cells with HIV-*gpt*(HXB-2) or HIV-I_{IIIB}*

Inhibitor concentration† (μM)	Percent inhibition‡ −1 hour HIV-gpt (HXB-2)	HIV-1_{IIIB}	+15 min HIV-1_{IIIB}	+30 min HIV-I_{IIIB}
UK-88,947				
6	15	34.5	25	1
12	34	48.8	35	5
24	81	88.1	72	8.9
IC_{50} (μM)	15.8	12.6	18	–
IC_{90} (μM)	25.5	24.6	–	–
Ro 31-8959				
1	45.2	18.6	17	1
5	58.3	54.3	48	8
10	75.2	88.6	80	13.2
20	90.1	99.7	80.8	13
IC_{50} (μM)	3.1	4.3	5.3	–
IC_{90} (μM)	12.8	10.2	–	–

*HIV-*gpt*(HXB-2) was produced with cotransfection of COS-7 cells with HIV-*gpt* plasmid DNA and HXB-2 *env* DNA (1.2×10^4 IU/ml). HIV-1_{IIIB} was cultured in H9 cells (2.1×10^5 IU/ml). HeLaCD4-LTR/β-gal cells were infected with 0.1 moi of each of the viruses for 1 hour. To prevent virus spread, rsCD4 (5 μg/ml) was added to the cultures 1 hour after infection with HIV-1_{IIIB}.

†Inhibitor was added to the cultures at the indicated time intervals before or after infection with virus. At 40 hours after infection, X-gal was added and blue foci of treated and untreated cultures were counted. The percentage of inhibition was calculated as the mean of blue cells in inhibitor-treated cultures compared to that in untreated cultures. IC values were determined by Hill equation analysis. When added at −1 hour, AZT inhibited HIV-1_{IIIB} infection with IC_{50} and IC_{90} values of 0.2 and 2.3 μM, respectively.

‡Stock solutions were made in DMSO. No cytotoxicity was seen even at the highest concentration of DMSO (0.2%) and inhibitors. TD_{50} was 72 μM for UK-88,947 and 40 μM for Ro 31-8959.

Proteolytic Cleavage of the EIAV NC Protein During Endogenous Reverse Transcription

We have previously reported that EDTA induced processing of the NC protein by the retroviral PR in cores (capsids) isolated from EIAV. From these data, we have proposed that in addition to its role in maturation, the PR may also function in the early phase of replication. We found that the NC protein in the EIAV cores was also cleaved by the PR during endogenous reverse transcription. When the EIAV cores were incubated at 37°C under conditions of endogenous reverse transcription, the PR cleaved the NC protein in each of the two conserved zinc fingers. Furthermore, as seen in the presence of EDTA, processing was ordered such that the first cleavage occurred between Cys24 and Tyr25 and the second cleavage occurred between Cys43 and Phe44.

The cDNA synthesized by reverse transcriptase (RT) was the 8.2-kilobase double-stranded EIAV cDNA. Interestingly, at some time points during reverse transcription, the migration of the 8.2-kilobase cDNA in a native agarose gel was retarded. The appearance of this gel-retarded cDNA correlated with the PR processing of the NC protein at the second cleavage site (Figure 7). The inclusion of a retroviral PR inhibitor in the endogenous RT reaction mixture not only inhibited processing of the NC protein, but blocked the appearance of the gel-retarded cDNA (Figure 8). These data are consistent with the involvement of the NC protein and PR in the early phase of the viral life cycle and suggest that PR cleavage of the NC protein may be necessary for the formation of a cDNA-protein complex (a preintegration complex). The characterization of the cDNA-protein complex is in progress.

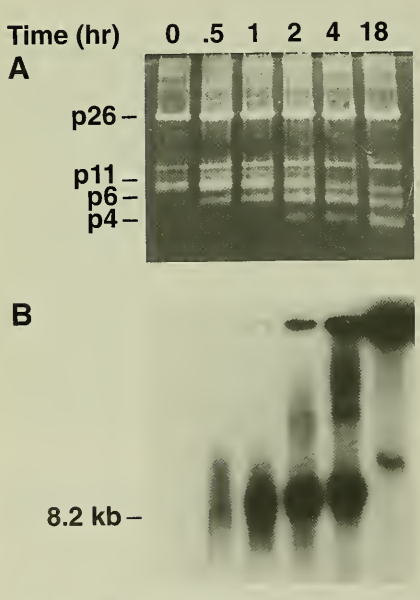


FIGURE 7. Correlation of EIAV NC protein processing with cDNA synthesis during endogenous reverse transcription. (A) SDS-PAGE analysis of viral proteins and NC protein cleavage products p6 and p4. (B) Agarose gel analysis and autoradiography of cDNA and the gel-retarded cDNA-protein complex.

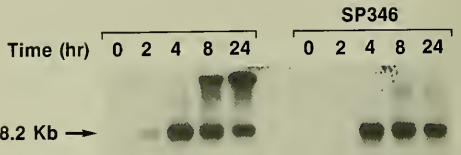


FIGURE 8. Endogenous reverse transcription of EIAV cDNA in the presence and absence of PR inhibitor (SP346). The DNA was analyzed by agarose gel electrophoresis.

Nucleolar Targeting Signal of HTLV-I Rex Protein Specifically Binds to Nucleolar Shuttle Protein B-23

Rex protein, the posttranscriptional regulator of HTLV-I, is located predominantly in the nucleolus and is associated with the cytoplasmic accumulation of unspliced and singly spliced viral mRNAs. The N-terminal 19-amino acid segment of Rex has been identified as the nucleolar targeting signal (NOS) and shown to be important for Rex function (Siomi et al., Cell 55:197, 1988). Last year, we reported studies on the molecular interaction between the NOS peptide and its binding host protein. A 38-kilodalton nucleolar protein was found to specifically interact with NOS. During the past year, we purified this protein by immunoaffinity chromatography and chemically characterized it. We initially attempted N-terminal sequencing; however, the results indicated a blocked N terminus. Therefore, we prepared chymotryptic peptides of the 38-kilodalton protein and determined their N-terminal sequence by Edman degradation. The results are shown in Figure 9. A database search indicated that the sequences of these peptides correspond to residues 34 to 42 (P-30), 48 to 62 (P-23), 50 to 67 (P-21), 137 to 151 (P-31), 218 to 228 (P-28), 252 to 266 (P-26), and 277 to 286 (P-22) of the 294-residue human nucleolar protein B-23. Moreover, the amino acid composition of the 38-kilodalton protein agreed well with that of B-23. These results clearly identify the major NOS-binding protein (38-kilodalton protein) as the nucleolar shuttle protein B-23.

To demonstrate the interaction between NOS and B-23 protein in more detail, we attempted to determine the region(s) of B-23 that binds to NOS. For these binding

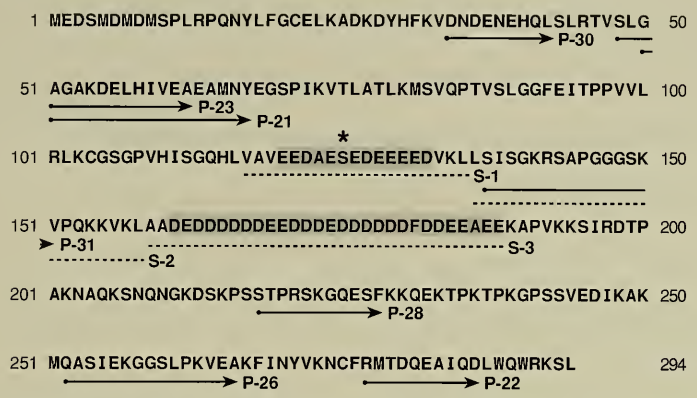


FIGURE 9. Identification of the 38-kilodalton NOS-binding protein as nucleolar protein B-23. Solid lines with arrowheads indicate the determined amino acid sequences of chymotryptic peptides from the 38-kilodalton nuclear protein bound to NOS peptide by Edman degradation. Dashed lines indicate amino acid sequences of the synthetic peptides. Shaded areas indicate highly acidic sequences. The asterisk indicates an in vivo phosphorylation site. All of these peptides were used for the binding study.

studies, we used the chymotryptic peptides of B-23 and synthetic peptides (Figure 9) and examined the binding ability of each peptide to immobilized NOS-W after ^{125}I-labeling. Two peptides, S-1 and S-3, exhibited the highest degree of binding to NOS. These results indicate that two regions of B-23 that contain long stretches of negatively charged amino acids, residues 117 to 135 and 159 to 188, are acceptor regions for NOS of Rex.

Rex protein, which plays a key role in the expression of HTLV-I structural proteins by the *gag* and *env* genes, is thought to mediate, via the RXRE, the transport of unspliced (*gag*) and singly spliced (*env*) mRNAs from the nucleus to the cytoplasm. Our results now suggest that B-23 functions not only as a shuttle protein for the nucleolar transport of ribosomal components but also as a shuttle for the import of Rex from the cytoplasm to the nucleolus coupled to the export of viral mRNAs containing the RXRE. These studies of Rex were carried out in collaboration with Yoshifumi Adachi (who is now in the Human Retrovirus Section, Molecular Mechanisms of Carcinogenesis Laboratory), Terry Copeland (Special Program in Protein Chemistry), and Masakazu Hatanaka (Institute for Virus Research, Kyoto University, Kyoto, Japan).

Expression and Purification of the MMTV dUTPase

dUTPase, an enzyme that hydrolyzes dUTP to dUMP and PP_i, has been found in prokaryotes and eukaryotes, as well as in certain retroviruses, herpesviruses, and poxviruses. In retroviruses, the dUTPase-encoding region is located either in the *pol* gene, between the genes for RT and integrase [e.g., in feline immunodeficiency virus (FIV)] or at the 5′end of the *pro* open reading frame (ORF), as in MMTV. The MMTV *gag/pro* ORF-encoded transframe protein (p30) contains the NC protein domain derived from the 3′ end of *gag*, fused to 154 amino acid residues encoded by the 5′ region of the *pro* ORF (Figure 10).

We cloned the DNA coding region for p30 into the plasmid pALTER-1 and inserted an additional nucleotide by site-directed mutagenesis to allow readthrough from the *gag* gene into the *pro* frame. We then cloned the insert into pGEX-2T. We isolated the chimeric protein (GST-p30) by affinity chromatography on a glutathione Sepharose 4B column, and after thrombin treatment, purified the p30 further by chromatography on a single-strand DNA-agarose column. This protein showed dUTPase activity, with only negligible cleavage of dATP, dCTP, dTTP, or UTP.

```
AAAMRGQKYSTFVKQTYGGGKGGQGAEGPVCFSCGKTGHIRKDCKDEKGS      50
                                                *
KRAPPGLCPRCKKGYHWKSECKSKFDKDGNPLPPLETNAENSKNLVKGQS     100

PSPAQKGDGVKGSGLNPEAPPFTIHDLPRGTPGSAGLDLSSQKDLILSLE     150

DGVSLVPTLVKGTLPEGTTGLIIGRSSNYKKGLEVLPGVIDSDFQGEIKV     200

MVKAAKNAVIIHKGERIAQLLLLPYLKLPNPVIKEERGSEGFGSTSHVH      249
```

FIGURE 10. Amino acid sequence of MMTV transframe protein p30, showing the NC protein (residues 1 to 95, indicated by the asterisk) derived from *gag* and the dUTPase domain (residues 96 to 249) derived from the *pro* ORF.

The enzyme was inhibited by EDTA, but its effect could be reversed by Mg^{2+}. We also detected dUTPase activity in purified MMTV, and p30 was the only protein recognized by antibodies directed towards the C-terminal sequence of the dUTPase coding region. We are conducting a further characterization of this interesting enzyme, in which the dUTPase sequence, as found in some retroviruses (e.g., FIV, EIAV), is fused to the NC protein.

Retroviral Genetics Section

Alan R. Rein

Ya-Xiong Feng
William Fu
Jacqueline Gordon
Demetria Harvin
Jane Mirro

The research of the Retroviral Genetics Section is directed at a molecular understanding of mechanisms of replication and pathogenesis by retroviruses. Work in the current year has focused on the packaging of dimeric viral RNA in retrovirus particles and on the mechanism of tumor induction by the 10A1 isolate of murine leukemia virus.

Maturation of the Dimeric Viral RNA of Moloney Murine Leukemia Virus

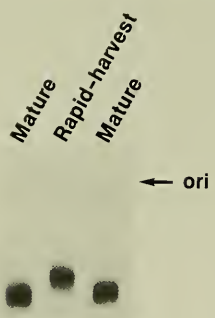
In all retrovirus particles, the genomic RNA is in the form of a dimer. The structure of the dimer is not fully understood: two monomers of the same polarity are evidently linked by hydrogen bonds or other weak, noncovalent bonds, since the dimer is dissociated into monomers under relatively mild conditions.

It is also not known where and when the dimer is formed. On one hand, some early studies reported that RNA isolated from rapid-harvest avian virus is apparently monomeric; these results strongly suggested that the dimer forms only after the virus is released from the cell. However, other studies have analyzed the RNAs in murine leukemia virus (MuLV) particles under conditions in which packaging of viral RNA is very inefficient, i.e., after treatment of virus-producing cells with actinomycin D, or in cysteine-array viral mutants. Under these conditions, the majority of released virions lack genomic RNA. Remarkably, the limited amount of RNA that is packaged under these conditions is still dimeric. This observation, in contrast to the studies on rapid-harvest virus discussed above, implies that the monomers of genomic RNA are not packaged independently of each other but, rather, are already joined into a dimeric structure when they are packaged.

One way to reconcile all of these observations is to suggest that the RNA is initially packaged as a dimer, but that this dimer is more fragile than that found in mature virions and is therefore dissociated under some extraction conditions. After the virus is released from the cell, the dimer undergoes a "maturation" or stabilization event.

To test this possibility, we extracted RNA from Moloney MuLV (Mo-MuLV) virions isolated every 5 minutes. Electrophoretic analysis of the RNA from these rapid-harvest virions (Figure 1) shows that dimeric RNA was obtained, but this RNA migrated more slowly than that isolated from control (24 hour) virus preparations. This observation supports the idea that the dimeric RNA undergoes some change in state after the virus is released from the cell.

We also tested the idea, discussed above, that the dimeric structure of the RNA in newly released virus is less stable than that in mature virus. RNA from rapid-harvest or mature virus preparations was incubated at different temperatures before electrophoretic analysis. In each case, the proportion of RNA that was still in the dimeric form was determined. As shown in Figure 2A, dimeric RNA from rapid-harvest virus dissociated into monomers at a slightly lower temperature than that from mature virus. The autoradiographs in Figure 2A were analyzed by densitometry; as shown in Figure 2B, the T_m (the temperature at which only half of the RNA remains in the dimeric form) of rapid-harvest viral RNA was about 51.3°C, while that of the mature viral RNA was about 53.8°C.



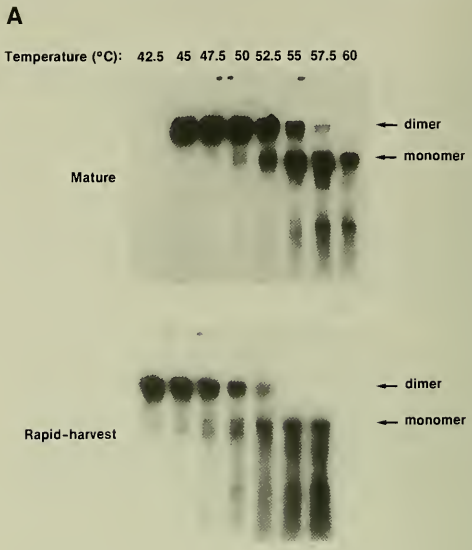


FIGURE 1. Analysis of viral RNA from rapid-harvest Mo-MuLV on a nondenaturing gel. RNAs from virions isolated after incubation for 5 minutes (rapid-harvest) and 24 hours (mature) were analyzed by the nondenaturing Northern blot procedure, using a Mo-MuLV riboprobe. Mature RNA was applied to two lanes to show the reproducibility of its migration under these conditions.

We reported last year that the dimeric RNA extracted from virions of protease (PR)-deficient mutants of MuLV also migrates more slowly in nondenaturing gel electrophoresis, and melts at a lower temperature, than that from mature wild-type virus. These observations suggested that PR function is required for some change in the conformation and stability of the dimeric RNA of the virus. However, an alternative interpretation is that the differences in the RNAs are due to sequence differences rather than the lack of PR function in the mutants. To test this possibility, we analyzed the effect of PR function on the mobility of dimeric RNA of Kirsten sarcoma virus (KiSV). KiSV is a replication-defective derivative of MuLV. KiSV RNA encodes no viral proteins but is efficiently packaged by MuLV proteins. As shown in Figure 3, KiSV dimeric RNA migrated more slowly in nondenaturing gel electrophoresis after packaging by PR^- viral proteins (ΔPR in lane 1, D32L in lane 5) than by wild-type proteins (lanes 2 and 4).

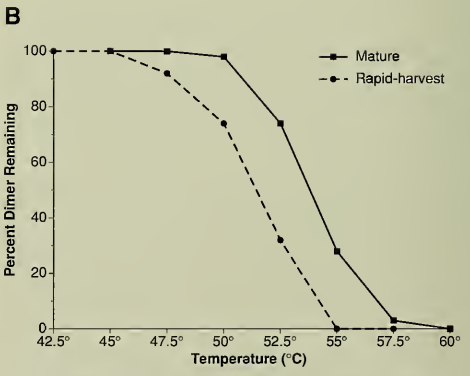


FIGURE 2. Melting curves of dimeric RNAs isolated from rapid-harvest and mature virions. (A) RNAs were treated at the indicated temperatures, and then analyzed by the nondenaturing Northern blot procedure using a Mo-MuLV riboprobe. (B) The data shown in panel A subjected to densitometry. Note that mature viral RNA was analyzed over the range 45°C to 60°C, while rapid-harvest viral RNA was treated with temperatures ranging from 42.5°C to 57.5°C.

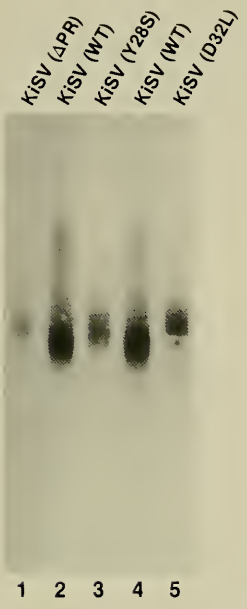


FIGURE 3. Electrophoresis of RNA dimers from KiSV. Dimeric RNAs of KiSV rescued by wild-type (lanes 2 and 4), ΔPR mutant (lane 1), D32L PR⁻ mutant (lane 5), and the cysteine-array mutant Y28S (lane 3) of Mo-MuLV were analyzed by the nondenaturing Northern blot procedure, using a c-Ki-*ras* probe to detect the KiSV genome.

The fact that the difference between wild-type and PR⁻ RNAs is seen with KiSV RNA, as well as with Mo-MuLV RNA itself, shows that PR acts in trans to affect the structure of dimeric viral RNA.

In general, mutants in the cysteine-array motif of the nucleocapsid (NC) region of the Gag polyprotein package very little intact viral RNA. However, at least some cysteine-array mutants of Mo-MuLV are only partially, rather than completely, defective with respect to packaging of KiSV RNA. This "leakiness" in the packaging of KiSV RNA allowed us to investigate the properties of viral RNA isolated from a cysteine-array mutant (Y28S), in which the Gag polyprotein is cleaved normally by the PR but the structure of the cysteine-array motif in the NC protein has been disrupted by mutation. We therefore isolated RNA from a virus preparation in which KiSV RNA had been packaged by the Y28S mutant of Mo-MuLV. As shown in Figure 3, lane 3, this dimeric RNA, like that packaged by PR⁻ Mo-MuLV, migrated more slowly in nondenaturing gel electrophoresis than KiSV RNA packaged by wild-type Mo-MuLV.

To explain these findings, we suggest that retrovirus particles undergo a multistep "maturation" pathway, as follows: (1) Gag polyprotein molecules assemble into a nascent virion containing an immature dimer; (2) the particle is released from the cell, and the Gag polyprotein is cleaved by PR, generating cleavage products including NC protein; and (3) NC protein, in a step requiring the cysteine array, acts on the dimer, converting it from immature to mature form.

One important implication of the ideas discussed here is the possibility that only dimeric RNAs are packaged in retroviruses. Although a number of reports have indicated that retroviruses contain a mixture of monomeric and dimeric RNAs under certain conditions, these conditions may all be cases in which the dimer is immature and more fragile than the mature wild-type dimer. Thus, it is conceivable that the monomers observed under these conditions actually arise by dissociation of fragile dimers during RNA extraction.

The idea that only dimeric RNAs are packaged also raises the possibility that the dimeric structure is an element of the "packaging signal" recognized by the Gag polyprotein during virus assembly. Thus, experiments designed to determine which viral sequences are necessary or sufficient for encapsidation of the RNA may, in fact, determine the sequence requirements for dimerization.

Induction of Early, Relatively Undifferentiated Hematopoietic Tumors by 10A1 MuLV

Different isolates of MuLV exhibit striking differences in their biological properties. One isolate, termed 10A1, has a number of unique characteristics. We have performed a partial characterization of the tumors induced in newborn NIH/Swiss mice by 10A1 MuLV. These mice develop leukemia with a mean latency of 120 days. The mice exhibit splenomegaly but do not show tumor involvement of the thymus or lymph nodes. These mice also have severely depressed hematocrits.

We have performed a histopathological characterization of 10A1-induced tumors by examining peripheral blood cells from tumor-bearing mice, in collaboration with Miriam Anver (Program Resources, Inc./DynCorp, NCI-FCRDC). These studies showed that two different morphologies are present. One is a blast-like cell that is negative for nonspecific esterase and hemoglobin. These cells appear to be extremely early erythroid precursors, i.e., proerythroblasts. The second cell type is an irregularly shaped cell with acentric, bean-shaped nuclei. These cells are also negative for nonspecific esterase and hemoglobin but, in contrast, have the characteristics of very early myeloid precursors, i.e., promyeloblasts.

In our earlier analysis of 10A1-induced tumor cells by fluorescence-activated cell sorting (FACS), these cells were negative (lin^-) for the lineage-specific markers CD11b, Mac-2, 8C5, Ia, CD32, CD45, CD4, CD8, Thy-1, CD72, and CD45R. We found that 10A1-induced tumor cells were positive only for the CD11a (LFA-1) marker. This marker is present on many different cell types, including undifferentiated precursors (Holmes and Morse, Immunol Today 9:344, 1988). Since histopathological analysis and the previous data indicate that these tumor cells are early undifferentiated cells, we next examined cells for the presence of markers found on either early lineage-restricted or multipotent cellular precursors.

Two markers, termed Sca-1 (Spangrude et al., Science 241:58, 1988) and Fall-3 (Müller-Sieburg, J Exp Med 174:161, 1991), are found on many differentiated cell types; however, when present on lin^-, Thy^{lo} cells, they are markers for multipotent precursors. Since 10A1-induced tumor cells are uniformly lin^- (though not Thy^{lo}), we tested these cells for Sca-1 and Fall-3 expression. Our FACS analysis indicated that six of seven tumors were Sca-1 positive (Table 1). Furthermore, 9 of 11 tumors were positive for Fall-3. Thus, a significant number of individual tumor cells have these markers for multipotent precursor cells.

Recently, a marker found only on stem cells, AA4.1, was described (Jordan et al., Cell 61:953, 1990). Our FACS experiments showed that 10A1-induced tumor cells were not positive for this stem cell marker (data not shown).

Since our earlier report suggested that these cells might be early erythroid precursors, markers that are expressed early in various lineages were also examined. One such marker, YW25.12.7, is expressed on myeloid and erythroid cells (Bertoncello et al., Exp Hematol 19:95, 1991). Six out of six tumors tested for this marker were positive (Table 1).

A lymphoid stem cell marker (Joro 30-8) along with two pro-T cell markers (Joro 75 and Joro 37-5) described by Palacios et al. (J Exp Med 172:219, 1990) were also tested by FACS analysis. Out of 19 tumors tested for the stem cell

TABLE 1. FACS analysis of tumor cells

	Marker*						
Tumor #	Sca-1	Fall-3	Joro 30-8	Thy-1	LFA-1	YW25.12.7	Lip-6
724927-1	+	–	–	–	+	+	ND
724927-3	+	+	–	–	+	+	ND
41592	+	+	ND	ND	ND	ND	ND
4792	–	+	ND	–	ND	+	ND
529927-2	+	+	ND	–	ND	ND	ND
529927-1	+	+	ND	ND	ND	ND	ND
529929-1	+	+	ND	ND	ND	ND	ND
716911	ND	ND	ND	–	+	SL	+
716912	ND	ND	ND	–	+	+	+
716913	ND	ND	ND	–	+	+	+
917918-3	ND	+	–	–	+	ND	ND
917914-6	ND	–	–	–	+	ND	ND
917918-9	ND	+	–	–	+	ND	ND
917918-7	ND	SL	–	–	+	ND	ND
J	ND	ND	–	–	+	ND	ND
K	ND	ND	–	–	+	ND	ND
L	ND	ND	–	–	+	ND	ND
H	ND	ND	SL	–	+	ND	ND
I	ND	ND	+	ND	+	ND	ND
A	ND	ND	–	ND	ND	ND	ND
AB	ND	ND	+	ND	ND	ND	ND
AC	ND	ND	+	ND	+	ND	ND
AD	ND	ND	–	ND	+	ND	ND
B	ND	ND	–	–	+	ND	ND
C	ND	ND	+	–	+	ND	ND
D	ND	ND	+	–	+	ND	ND
E	ND	ND	+	–	+	ND	ND
Total	6/7	9/11	7/19	0/19	20/20	6/6	3/3

*Symbols are: +, positive for fluorescence; –, negative for fluorescence; SL, slightly positive for fluorescence; ND, not determined.

marker Joro 30-8, seven were positive. In contrast, the cells were negative for the Joro 75 and Joro 37-5 pro-T cell markers. Finally, three of three tumors were positive for another early lymphoid cell marker, the Lip-6 marker (described by Holmes et al., personal communication).

We previously reported that 10A1-induced tumors do not show rearrangements in either their T cell receptor or immunoglobulin genes, which provides strong evidence that the tumor cells are not differentiated cells in the lymphoid pathway. We have recently extended this analysis by examining the tumor cells for expression of certain genes characteristic of hematopoietic progenitors. Figure 4 shows a Northern blot analysis of a series of tumor samples. As shown in Figure 4A, nearly all of the tumors expressed T cell receptor-γ. While expression of this gene is normally restricted to lymphoid cells, the fact that the locus is not

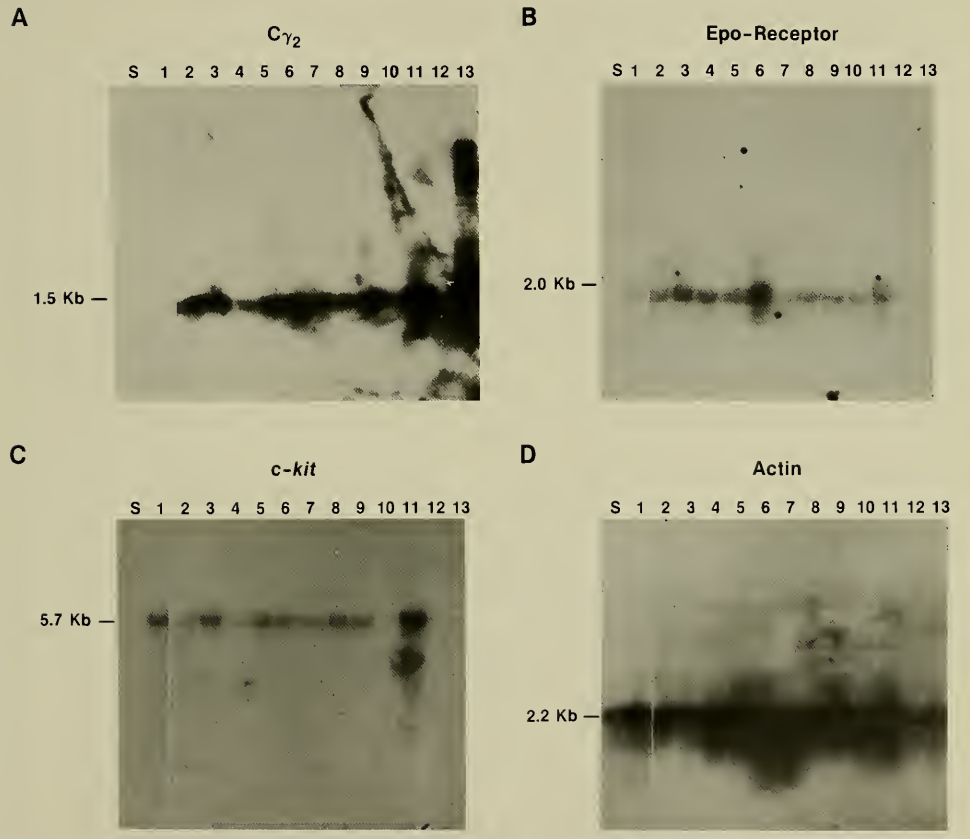


FIGURE 4. Northern blot analysis of total RNA from leukemic spleens. A Northern blot of total RNAs from MuLV-induced tumors was hybridized with the following probes: (A) T cell receptor-γ cDNA clone; (B) erythropoietin-receptor cDNA clone; (C) *c-kit* cDNA clone; (D) human β-actin cDNA clone. Lane S, total RNA from spleen; lanes 1 to 12, total RNA from 10A1-induced tumors; lane 13, total RNA from a Mo-MuLV–induced tumor.

rearranged in these tumors indicates that the transcripts detected here are probably "sterile" transcripts. The expression of noncoding transcripts of constant-region genes has previously been observed very early in prolymphocyte differentiation (e.g., Weinstein et al., J Exp Med 169:2059, 1989).

When the same Northern blot was stripped and probed with an erythropoietin receptor (EpoR) probe, as seen in Figure 4B, most tumors (10 out of 12) expressed the EpoR. The Epo receptor is present not only on differentiating erythrocytic cells but on earlier, multipotent cells as well (Heberlein et al., Mol Cell Biol 12:1815, 1992). Despite the presence of this RNA, we have found that the cells are virtually unresponsive to erythropoietin (data not shown).

We also tested the same blot with a c-*kit* probe. As shown in Figure 4C, 11 out of the 12 10A1-induced tumors expressed this gene. This gene is expressed in hematopoietic stem cells and encodes the receptor for stem cell growth factor (reviewed by Witte, Cell 63:5, 1990).

In summary, the gene expression pattern that we have observed in the 10A1-induced tumor cells, like their phenotype in FACS analysis, is not diagnostic of any single hematopoietic lineage but is consistent with the idea that the cells are very early precursor cells.

In many cases, MuLV leukemogenesis is correlated with retroviral integration and activation of proto-oncogenes. We used Southern blot analysis to determine whether two loci, c-*myc* and *fli-1*, which are frequent targets of proviral integration, were rearranged in 10A1-induced tumors. Based on our analysis, the c-*myc* locus was not rearranged in 12 out of 12 tumors (data not shown). However, Southern blot analysis demonstrated that 14 out of 14 10A1-induced tumors contained a similar-sized rearrangement of the *fli-1* gene (Figure 5). Southern blot analysis also showed that all these integrations were within 200 base pairs of the normal translational initiation site of *fli-1*. 10A1-induced tumors, unlike other MuLV-induced tumors, had acquired a comparatively small number of proviruses (one to six), as determined by "unblot" analysis with a 10A1-specific oligonucleotide; these *fli-1* integrations were detected as a common site of proviral integration by this analysis (data not shown). To determine the structure of the 5′ region of *fli-1* mRNAs in 10A1-induced tumor cells, we used reverse transcriptase-polymerase chain reaction (RT-PCR) with 10A1- and *fli-1*–specific primers to clone and analyze the 10A1–*fli-1* hybrid transcript in three 10A1-induced tumors. Cloning and sequencing of the RT-PCR products revealed, as indicated in Figure 6, that in all three cases the 10A1

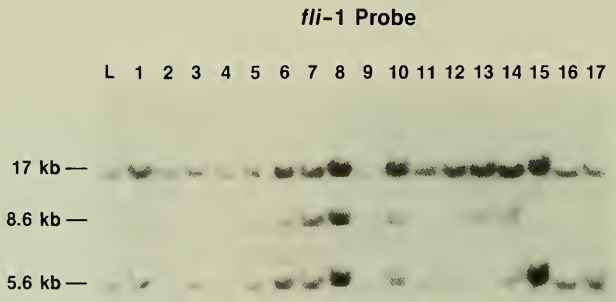


FIGURE 5. Rearrangement of the *fli-1* locus in 10A1-induced tumors. Cellular DNA was extracted from tumors, digested with *Eco*RI, and probed with an *fli-1* probe. Lanes: L, normal liver; 1 to 14, 10A1-induced tumors; 15, a Mo(4070A)-induced tumor; 16 and 17, Mo-MuLV–induced tumors.

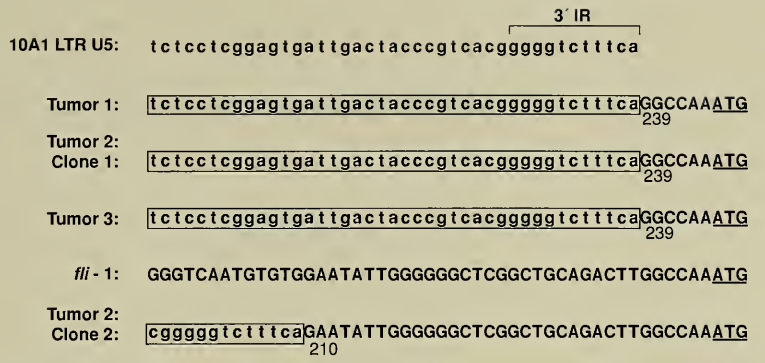


FIGURE 6. 10A1 MuLV–*fli-1* junctions in chimeric RNAs isolated from 10A1-induced tumors. RT-PCR was performed on RNA isolated from these tumors, using a (+)-strand primer from 10A1 and a (–)-strand primer from *fli-1*. Products of this amplification reaction were cloned and sequenced; the figure shows the junction region in four of the resulting clones. IR, inverted repeat region of viral long terminal repeat.

3′ long terminal repeat sequences were joined to the *fli-1* 5′ untranslated region at nucleotide 239. Interestingly, one tumor had two 10A1–*fli-1* hybrid mRNAs. In addition to the aforementioned hybrid, a 10A1 MuLV integration also occurred at nucleotide 210 in this tumor.

Integrations at nucleotide 239 of *fli-1* have previously been described in the characterization of tumors induced by Cas-Br MuLV (Bergeron et al., Virology 191:661, 1992). These tumors have been referred to as "non-T, non-B lymphomas" and resemble 10A1-induced tumors in their lack of obvious lineage-specific markers (Bergeron et al., J Virol 65:7, 1991).

Fli-1, a member of the *ets* gene family, was originally identified as a proviral integration site in tumors induced by Friend mink cell focus-forming (MCF) virus and Friend MuLV (Ben-David et al., Proc Natl Acad Sci USA 87:1332, 1990). These tumors were clearly identified as erythroleukemias. It is intriguing to note that in these tumors, the provirus was integrated approximately 2 kilobases 5′ of the transcriptional initiation site of *fli-1* and in the opposite transcriptional orientation (Ben-David et al., Genes Dev 5:908, 1991). These results suggest that in these erythroid tumors, the provirus was contributing to malignant transformation by means of enhancer insertion. In contrast, in the less-differentiated tumors induced by 10A1 MuLV and Cas-Br MuLV, the transcriptional activation of *fli-1* seems to result from a promoter-insertion mechanism. Perhaps these differences in activation mechanism are responsible for the variation in the target cells that has been observed with the different MuLV isolates.

Molecular Biology of Retroviruses Section

Nancy R. Rice

Linda L. Dunn
Mary K. Ernst
Tadeusz J. Kochel
Lyudmila A. Lyakh

Studies in the Molecular Biology of Retroviruses Section focus on the *rel* oncogene and Rel protein, and on other members of the nuclear factor-κB (NF-κB) family of transcription factors. During the current year, we investigated the most prominent inhibitor of these proteins—IκBα—and followed the release of active NF-κB upon stimulation of cells.

NF-κB is a transcription factor that plays a crucial role in the regulation of numerous genes, including many genes involved in immune or inflammatory responses. NF-κB was originally described as a heterodimer composed of a p50 and p65 subunit. However, it is now known that there are five different proteins that can participate in NF-κB dimer formation: p52, c-Rel, and Rel B (also called I-Rel), in addition to p50 and p65. Over a region of about 300 amino acids, all of these proteins are related to the c-Rel oncogene product and to the *Drosophila* morphogen Dorsal. DNA-binding forms of these proteins can be homodimers (except for Rel B) or many (or perhaps all) of the possible heterodimers. Transcriptional activation is due chiefly to p65, c-Rel, and Rel B.

NF-κB activity is constitutive in B cells and in some monocyte cell lines, but in most other cells it is very low or undetectable. However, NF-κB activity can be rapidly induced in most cells by treatment with any of a variety of agents, e.g., phorbol esters, tumor necrosis factor (TNF), or hydrogen peroxide. The fact that in most cases this induction does not require protein synthesis led to the discovery that NF-κB is stored in inactive form in the cytoplasm. Storage of c-Rel- and p65-containing dimers is accomplished through binding to one of several inhibitor proteins.

The most extensively studied inhibitor of the Rel protein family is IκBα, the product of the MAD-3 gene. In vitro, IκBα binds efficiently to c-Rel, p65, and Rel B and prevents DNA binding by their homo- or heterodimers. Cytoplasmic IκBα complexes are thought to be major targets in signal transduction pathways. It is known that phosphorylation of IκBα in vitro prevents its binding to c-Rel and to p50-p65, and it is presumed that phosphorylation regulates association in vivo as well. Whatever the precise mechanism, the working hypothesis is that stimulation of the cell results in dissociation of IκBα from cytoplasmic complexes and transport of Rel proteins to the nucleus, followed by DNA binding and gene activation.

In the past year, we examined IκBα-containing complexes directly by immunoprecipitation. We found that lipopolysaccharide (LPS) stimulation of murine B and pre-B cells results in dissociation of IκBα from all preexisting c-Rel- and p65-containing dimers, in accordance with the working model. Some but not all of the liberated c-Rel and p65 translocates to the nucleus and exhibits DNA-binding activity. The remaining c-Rel and p65 are captured by newly synthesized IκBα and retained in the cytoplasm, which strongly suggests that complexes dissociate because of modification to IκBα, not to c-Rel or p65. We also examined IκBα complexes in unstimulated cells and found a surprising dynamism. While c-Rel and p65 are stable proteins, complexed IκBα turns over with a half-life of 30 to 60 minutes in a variety of cell lines. We propose that this turnover results in the low level of active NF-κB that is presumably necessary even in the unstimulated cell. In addition, we suggest that the high rate of synthesis

of IκBα compared to that of Rel proteins provides the ability to turn off NF-κB activity rapidly as soon as the activating signal ceases.

Activation of Cells Results in Dissociation of IκBα

To test the effect of cellular stimulation on IκBα, we determined the response of the murine B cell line WEHI 231 and of the murine pre-B cell line 70Z/3 to bacterial LPS, an effective activator of these cells. Cells were incubated in medium containing [^{35}S]methionine and [^{35}S]cysteine for 1 hour, then in nonradioactive medium with or without LPS. We tested whether ^{35}S-labeled IκBα would dissociate from Rel complexes during incubation of the cells in LPS.

In WEHI 231 cells that were labeled and chased for 90 minutes without LPS, anti–c-Rel precipitated c-Rel, p105 (the precursor of p50, which also binds c-Rel and retains it in the cytoplasm), and IκBα (Figure 1, lane 1).

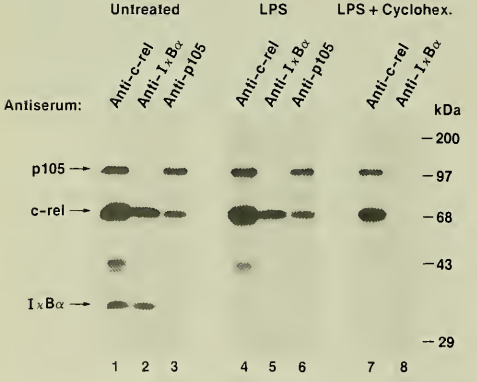


FIGURE 1. LPS results in dissociation of IκBα. WEHI 231 cells were incubated for 1 hour in medium containing [^{35}S]methionine and [^{35}S]cysteine followed by a chase period in nonradioactive medium containing LPS, LPS plus cycloheximide, or no additives. After 90 minutes, cytoplasmic extracts were immunoprecipitated with anti-c-Rel. The precipitates were eluted with excess peptide against which the antiserum was raised. The eluates were then reprecipitated with a second anti-c-Rel, anti-IκBα, or anti-p105 antiserum, as indicated, and analyzed by SDS-polyacrylamide gel electrophoresis (SDS-PAGE). Molecular size markers (in kilodaltons) are shown on the right.

Anti-IκBα precipitated c-Rel and IκBα (lane 2), and anti-p105 precipitated c-Rel and p105 (lane 3), showing that these are separate inhibitor complexes. In LPS-treated cells, the results were different. Anti–c-Rel precipitated c-Rel and p105, as before, but not IκBα (lane 4). Thus, in these B cells, LPS resulted in dissociation of ^{35}S-labeled IκBα from c-Rel–containing complexes. In contrast, p105 did not dissociate from c-Rel complexes and remained precipitable by both anti-c-Rel (lane 4) and anti-p105 (lane 6). Therefore, LPS affects preexisting IκBα complexes but not p105 complexes.

What happened to the c-Rel that was bound to ^{35}S-labeled IκBα? A small amount of the c-Rel entered the nucleus, where it was detectable by immunoblotting (data not shown), but most of it remained in the cytoplasm. In fact, in spite of the loss of ^{35}S-labeled IκBα, most of the ^{35}S-labeled c-Rel remained precipitable by anti-IκBα (Figure 1, lane 5). Our interpretation of this result is that displaced ^{35}S-labeled IκBα was rapidly replaced by newly synthesized (and, hence, nonradioactive) IκBα. If this conclusion were true, treatment with LPS in the presence of cycloheximide would prevent new synthesis and eliminate precipitability of c-Rel by anti-IκBα, which is exactly what happened (Figure 1, lane 8).

This experiment demonstrated that LPS treatment of B cells caused the loss of ^{35}S-IκBα from c-Rel complexes. Nevertheless, synthesis of IκBα continued in the presence of LPS, and newly synthesized molecules were able to associate with preexisting but newly liberated ^{35}S–c-Rel. These results strongly suggest that (1) the LPS signal affects IκBα complexes rather than free IκBα, i.e., the signal causes dissociation rather than preventing association, and (2) dissociation results from some change to IκBα, not to c-Rel, since ^{35}S–c-Rel remained able to bind to new inhibitor molecules. In addition, this experiment shows that translocation to the nucleus is not a necessary consequence of every dissociation event. If new IκBα is synthesized at a sufficiently high rate (others have shown that 90 minutes of exposure to at least some activating agents is sufficient to increase the level of IκBα mRNA significantly), then dissociated c-Rel and/or p65 may be efficiently recaptured and retained in the cytoplasm.

We wondered whether the ^{35}S-IκBα that dissociates from c-Rel complexes in response to LPS could be found as free molecules. We tested this possibility by exhaustively

precipitating the cytoplasmic extract from LPS-treated WEHI 231 cells with anti–c-Rel and anti-p65, then looking for ^{35}S-IκBα in the supernatant. ^{35}S-IκBα was not detectable (data not shown), suggesting that it is degraded soon after release.

The same experiments were performed with 70Z/3 pre-B cells, which are also activated by LPS. The results were the same as with the B cells except that a greater fraction of the liberated c-Rel and p65 was transported to the nucleus, where they were detectable by immunoblotting (data not shown) and by DNA-binding assays (see below). One obvious difference between the B and pre-B cells is the high level of Rel proteins and IκBα in the former. It remains to be seen whether this difference, and the attendant difference in rates of synthesis of these proteins, contribute to the efficient recapture of liberated c-Rel in the B cells.

DNA-binding assays revealed another difference between the pre-B and B cells. It has been known for some time that unstimulated pre-B cells have little or no active (i.e., nuclear) NF-κB, whereas unstimulated B cells have a lot. We found that there is a qualitative difference as well. The κB-binding activity in B cells, whether LPS-stimulated or not, was due largely to p50 and c-Rel, with little or no p65 involvement (Figure 2A). In contrast, κB-binding activity in LPS-stimulated pre-B cells was due to p50 and p65, with little c-Rel participation (Figure 2B). While this difference reflects the high level of c-Rel in B cells, its cause is unknown.

IκBα Has a Short Half-Life

The preceding experiments showed that disruption of IκBα complexes occurred upon activation of the cell. We wondered whether dissociation also occurs at a lower but significant rate in the unstimulated cell. This question was prompted by observations made by others that in pre-B cells, cycloheximide alone results in both nuclear κB binding activity and transcription of the kappa light chain gene. Both of these studies suggest that there is a short-lived protein inhibitor of kappa gene expression. If the labile protein is IκBα, then dissociation of IκBα complexes may be a frequent occurrence and not limited to periods of cellular activation.

To investigate the stability of IκBα, we incubated WEHI 231 cells in medium containing [^{35}S]methionine and [^{35}S]cysteine for 1 hour and then in nonradioactive

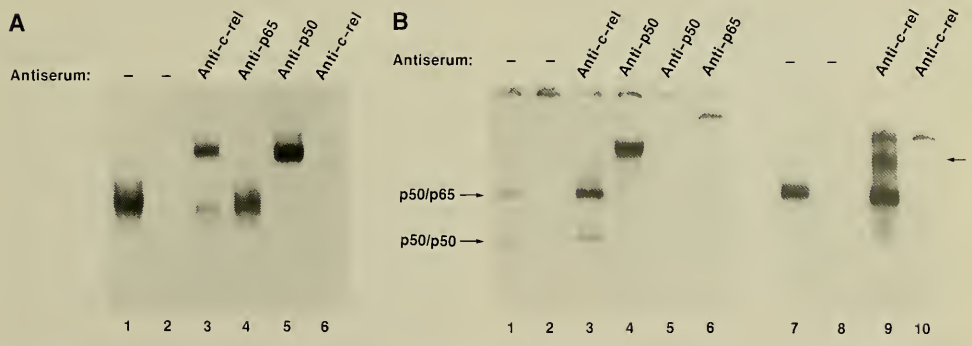


FIGURE 2. DNA-binding activity in B cells and in LPS-stimulated pre-B cells. (A) The nuclear extract from WEHI 231 cells was assayed for DNA-binding activity using a ^{32}P-labeled oligonucleotide containing the IL-6 κB site. Prior to addition of the ^{32}P-probe, extracts were incubated with anti-c-Rel (lane 3), anti-p65 (lane 4), or anti-p50 (lane 5). The sample in lane 6 contained the ^{32}P-probe and anti-c-Rel, but no nuclear extract. In lane 2, the incubation mixture included a 200-fold molar excess of unlabeled κB oligonucleotide. (B) 70Z/3 cells were treated with LPS, and nuclear extracts were assayed for the ability to bind to a ^{32}P-labeled oligonucleotide containing the murine kappa chain intronic κB site. Prior to incubation with the ^{32}P-probe, samples were incubated with two different anti–c-Rel sera (lanes 3 and 9), anti-p50 (lane 4), or anti-p65 (lane 6). Samples in lanes 5 and 10 contained the ^{32}P-probe and antiserum, but no nuclear extract. In lanes 2 and 8, the incubation mixture included a 40-fold molar excess of unlabeled κB oligonucleotide.

medium for various times. After boiling the total cell extract in SDS and precipitating it with anti-IκBα, we analyzed the immunoprecipitate by SDS-PAGE. The intensely labeled protein band seen in the pulse-labeled sample (Figure 3, lane 1) was nearly gone after 2 hours of chase (lane 3), and it continued to decline thereafter (lanes 4 to 6). The same result was obtained with three different IκBα antibodies, directed at different regions of the protein (data not shown). This result indicates that in these murine B cells, total cellular IκBα turns over very rapidly. From multiple experiments, we estimate a half-life of about 30 minutes.

To determine whether this result is unique to B cells, in which NF-κB is constitutively active, we performed similar experiments with the murine pre-B cell line 70Z/3, the human plasmacytoma cell line Sultan, the human T cell line Jurkat, and the human epithelial cell line HeLa. The results showed that IκBα turns over rapidly in these cells as well. In all cases, the half-life of IκBα was about 1 hour (data not shown).

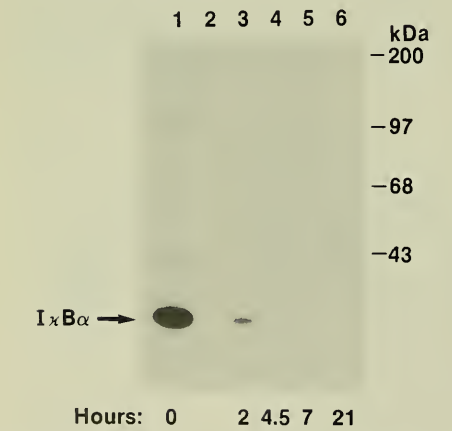


FIGURE 3. IκBα has a short half-life. WEHI 231 cells were incubated for 1 hour in medium containing [^{35}S]methionine and [^{35}S]cysteine, followed by a chase period in nonradioactive medium for the indicated times. Whole cell extracts were adjusted to 1% SDS and 0.5% β-mercaptoethanol and boiled for 5 minutes. After at least a fourfold dilution, aliquots were immunoprecipitated with anti-IκBα. The sample shown in lane 2 contained competing peptide. Molecular size markers are shown on the right.

Most IκBα Is Complexed to Rel and/or p65

The preceding experiments measured the stability of total cellular IκBα. These experiments leave open the possibility that there is a pool of free IκBα that is rapidly turning over, while complexed IκBα is quite stable. To determine the relative amounts of free versus complexed IκBα, we examined 70Z/3 and Jurkat lysates from which all c-Rel and p65 had been removed. ^{35}S-labeled cytoplasmic extracts were exhaustively immunoprecipitated with anti–c-Rel and anti-p65, and the amount of precipitable ^{35}S-IκBα in the supernatant was compared with that in nondepleted lysates. The result was that most IκBα was bound to c-Rel or p65; treating the lysate with anti–c-Rel and anti-p65 removed about 85% of the total IκBα (data not shown). We regard this figure as a minimum estimate of bound IκBα, since others have shown that additional molecules may be bound to p50 homodimers, which were not removed in these experiments. In fact, in a similar study with WEHI 231 cells, we found that the small amount of IκBα not precipitable by anti–c-Rel and anti-p65 was precipitated by anti-p50. Thus, in pre-B, T, and B cell lines, almost all IκBα is complexed rather than free.

Complexed IκBα Also Turns Over Rapidly

The observation that most IκBα is bound and that total IκBα has a short half-life necessarily implies that bound IκBα turns over rapidly. To test this prediction, we examined a specific subset of c-Rel–containing complexes in a pulse-chase experiment. WEHI 231 cells were labeled, chased for various times, and immunoprecipitated with anti-c-Rel. The washed precipitates were released from antibody by incubation in excess competing peptide, and the eluted (c-Rel–containing) material was reprecipitated with anti-p65 and analyzed by SDS-PAGE. Since each sample was precipitated sequentially with anti–c-Rel and anti-p65, each lane of the resulting gel shows complexes that contain both c-Rel and p65 (Figure 4A). In these complexes both ^{35}S–c-Rel and ^{35}S-p65 were quite stable over the course of 23 hours, suggesting that they associated with each other at or soon after synthesis and that they remained associated throughout the experiment. In contrast, ^{35}S-IκBα declined rapidly. The intense band seen in the pulse-labeled sample (Figure 4A, lane 1) was considerably reduced after

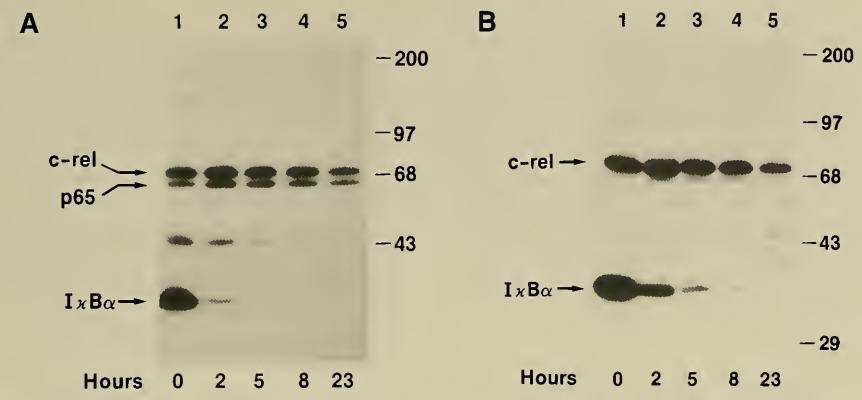


FIGURE 4. Complexed IκBα turns over rapidly. WEHI 231 cells were incubated in [^{35}S]methionine and [^{35}S]cysteine, followed by a chase period in nonradioactive medium for the indicated times. Cytoplasmic extracts were immunoprecipitated with anti-c-Rel, and precipitates were washed and eluted from antibody by incubation with excess competing peptide. Eluates were reprecipitated with anti-p65 (A) or anti-IκBα (B) and analyzed by SDS-PAGE. Molecular size markers are shown on the right.

a 2-hour chase (lane 2) and was almost gone after 5 hours (lane 3). Thus, the association between newly synthesized IκBα and the c-Rel–p65 complex was not stable, for the ^{35}S-IκBα exhibited a short half-life.

Next, we tested whether the displaced ^{35}S-IκBα was replaced with new IκBα molecules. As in the preceding experiment, c-Rel–containing complexes were collected at each time point. These complexes were reprecipitated with anti-IκBα and analyzed by SDS-PAGE. Thus, the gel shows only those complexes that contained both c-Rel and IκBα (Figure 4B). As in Figure 4A, pulse-labeled ^{35}S–c-Rel was stable, while ^{35}S-IκBα declined rapidly. However, the fact that anti-IκBα precipitated c-Rel even after ^{35}S-IκBα had almost disappeared (Figure 4B, lanes 4 and 5) shows that new unlabeled IκBα replaced the dissociated labeled molecules. This conclusion was confirmed by using cycloheximide in the chase period. The result was that after a 90-minute chase, anti-IκBα was unable to precipitate complexes containing ^{35}S-labeled Rel (data not shown). Thus, complexed IκBα is unstable. Even in unstimulated B cells, it frequently dissociates from c-Rel- and p65-containing complexes and is replaced by a newer molecule.

The rapid turnover of bound IκBα is not unique to B cells. In a pulse-chase experiment with 70Z/3 pre-B cells, we examined complexes containing IκBα and c-Rel or p65. The result was qualitatively the same as with B cells. A constant level of ^{35}S-labeled c-Rel or p65 was associated with IκBα, but ^{35}S-IκBα exhibited a half-life of about I hour. Thus, even in non-B cells, IκBα complexes are dynamic. Newly synthesized IκBα remains associated with Rel or p65 for a relatively short time, dissociates, and is replaced by a newer molecule. Since the dissociation rate mirrors the turnover rate of total IκBα, the dissociated molecule is most likely degraded.

The cost of maintaining this system is considerable. A 1-hour half-life for IκBα means that it will be replaced in half of all the complexes every hour, even in the absence of an activating signal. Viewed another way, if the half-life of IκBα-bound c-Rel and p65 is about 8 hours, IκBα is produced at 8 times the rate (at least) of their synthesis. What is the advantage to the cell of such a seemingly wasteful system? One possibility is that this arrangement generates the low level of NF-κB activity that is presumably necessary even in unstimulated cells. At a minimum, dividing cells must double their NF-κB content once per cycle, and active NF-κB is probably necessary to accomplish this (we and others have shown that the genes for both p105 and c-Rel have κB sites in their promoter regions). The continual turnover of IκBα may provide the opportunity for small amounts of liberated c-Rel/p65/p50 to

enter the nucleus. In contrast, if the complexes were very stable in the absence of an activation signal, there would be no active NF-κB for "housekeeping" purposes.

A second possible advantage of the system follows from the high rate of IκBα synthesis, which may provide the ability to curtail NF-κB activity rapidly once the activating signal subsides. It has been suggested by others that IκBα not only is the major cytoplasmic inhibitor of NF-κB, but also is responsible for turning off active NF-κB in the nucleus. The hypothesis is that some excess IκBα molecules may enter the nucleus, bind to active NF-κB, and release it from DNA. However, there can only be excess IκBα if its rate of synthesis exceeds that of the c-Rel and p65 to which it binds. If their rates of synthesis were the same, all new IκBα would bind to new c-Rel/p65: There would never be uncomplexed inhibitor, even in the presence of an activating agent. Thus, viewing this regulatory system as a whole, storage of active transcription factor in the cytoplasm enables instantaneous gene activation in response to cell stimulation. Constitutive synthesis of excess inhibitor may enable immediate damping of activity upon cessation or diminution of stimulation.

Mammalian Genetics Laboratory

Neal G. Copeland

Nancy A. Jenkins
Luis F. Parada
Peter J. Donovan

Research in the Mammalian Genetics Laboratory is focused on the identification and study of genes important in normal mammalian development and oncogenesis, using mice as a model system. The laboratory is currently composed of three sections and one working group: the Molecular Genetics of Oncogenesis Section, headed by Neal Copeland; the Molecular Genetics of Development Section, headed by Nancy Jenkins; the Molecular Embryology Section, head by Luis Parada; and the Cell Biology of Development and Differentiation Group, headed by Peter Donovan.

Inbred mouse strains provide important model systems for the identification of cellular genes causally associated with the disease process. The research staff of the Molecular Genetics of Oncogenesis Section has utilized this genetic resource to identify a number of novel cellular proto-oncogenes involved in murine hematopoietic disease. Members of this section are also studying a number of developmentally important genes that are potentially involved in human genetic disease.

The focus of the Molecular Genetics of Development Section is to identify and characterize genes important in normal mammalian development. Members of the Molecular Genetics of Oncogenesis and Molecular Genetics of Development Sections are also collaborating on a number of different research projects. One project involves the molecular genetic characterization of the mouse steel locus that has recently been shown to encode the ligand for the c-*kit* proto-oncogene product, a tyrosine kinase receptor. Work on the mouse steel locus is also being performed in collaboration with members of the Cell Biology of Development and Differentiation Group. A second project involves the development of a molecular genetic linkage map of the mouse genome using inter-specific mouse backcrosses.

The Molecular Embryology Section is studying the role of proto-oncogenes in mouse embryonic development. This section is exploiting the method of homologous recombination in embryonic stem cells, among other techniques, with the aim of elucidating normal function of proto-oncogenes as well as their function in tumorigenesis.

The Cell Biology of Development and Differentiation Group is studying the development of the germ line in the mouse embryo. Steel factor in combination with other growth factors such as basic fibroblast growth factor and leukemia inhibitory factor can promote long-term proliferation of germ cells in culture. The role of these factors in regulating germ cell survival, proliferation, and differentiation is being investigated. Other studies are aimed at understanding the potential role of cell cycle regulators in germ cell development, particularly entry into and regulation of meiosis.

Molecular Genetics of Oncogenesis Section

Neal G. Copeland

Mary Barnstead
Camilynn Brannan
Jeffrey Ceci
Linda Cleveland
David Largaespada
Xiaobei Liao
Takuro Nakamura
Norene O'Sullivan
Bhavani Pathak
Lisa Secrest
John Shaughnessy

Members of the Molecular Genetics of Oncogenesis Section are using mice as a model system to identify and clone novel cellular genes causally associated with murine disease and development. This year, in collaboration with the Molecular Genetics of Development Section and Colin Hodgkinson, Atsuo Nakayama, and Heinz Arnheiter (National Institute of Neurological Disorders and Stroke, Bethesda, MD), part of our work has focused on the analysis of the microphthalmia locus.

Mutations at the Mouse *microphthalmia* Locus Are Associated with Defects in a Gene Encoding a Novel bHLH-Zip Protein

Much of our knowledge on the function of genes in mammalian development and adulthood has come from the molecular analysis of spontaneous or induced mutations. One of the most interesting mutations in the mouse is *microphthalmia* (*mi*). Many different *mi* alleles have been described and some of them show complex interallelic interactions. The original *mi* allele is semidominant and was found in descendants of an irradiated male. Heterozygotes often have white spotting on the belly, head, and tail and less pigment in the iris than normal. Homozygotes have small, unpigmented eyes and lack melanocytes in the skin and inner ear; they are white, microphthalmic, and deaf. They are also deficient in mast cells and have an intrinsic defect in osteoblasts that leads to osteopetrosis and failure of the incisors to erupt. Other *mi* alleles share with the original *mi* allele a lack of or abnormality in melanocytes. However, the severity of the phenotypes varies considerably from allele to allele: *microphthalmia*-Oak Ridge (Mi^{or}), a semidominant allele, displays most of the defects seen with the original *mi* allele; red-eyed white (mi^{rw}), a recessive allele, shows only a loss of eye and coat pigment and reduced eye size; and *microphthalmia* vitiligo (mi^{vit}) produces a progressive depigmentation of the pelage, skin, and eyes, but no microphthalmia.

Interallelic complementation at the *mi* locus has been observed in heterozygous combinations of several alleles. For instance, the compound heterozygotes Mi^{wh} (*microphthalmia* white)/*mi* and Mi^{wh}/mi^{ws} (white spotted) have less severely affected eyes and more pigment than either homozygous Mi^{wh}/Mi^{wh} or mi^{ws}/mi^{ws} mice. Other allelic combinations do not show complementation but do show enhancement of the phenotype associated with a particular allele. An example of such an enhancement is seen with mi^{sp} (spotted): even when homozygous, mi^{sp} does not show a visible phenotype, but in combination with other *mi* alleles, a phenotype becomes manifest. These interallelic interactions suggest that the *mi* gene product may function as a protein dimer or in some other, higher ordered protein structure. Alternatively, the *mi* gene product is a multifunctional protein, or the *mi* locus contains two or more tightly linked genes.

Using a transgenic insertional mutation at *mi* that was discovered in transgenic mice originally generated to study the regulation of the vasopressin promoter, we have cloned and characterized the *mi* gene.

Cloning the Regions Flanking a Transgene Insertion at *mi*

Pronuclear injection of a linear 6.2-kilobase fragment (Figure 1A) resulted in the generation of the VGA-9 mouse line, which contained approximately 50 copies of the transgene integrated at a single site. Mice bred to homozygosity for the transgene insertion were white and microphthalmic and showed early-onset deafness. Their phenotypic similarity to mice homozygous for mutations at *mi* and subsequent allelism studies suggested that the transgene insertion had occurred at the *mi* locus.

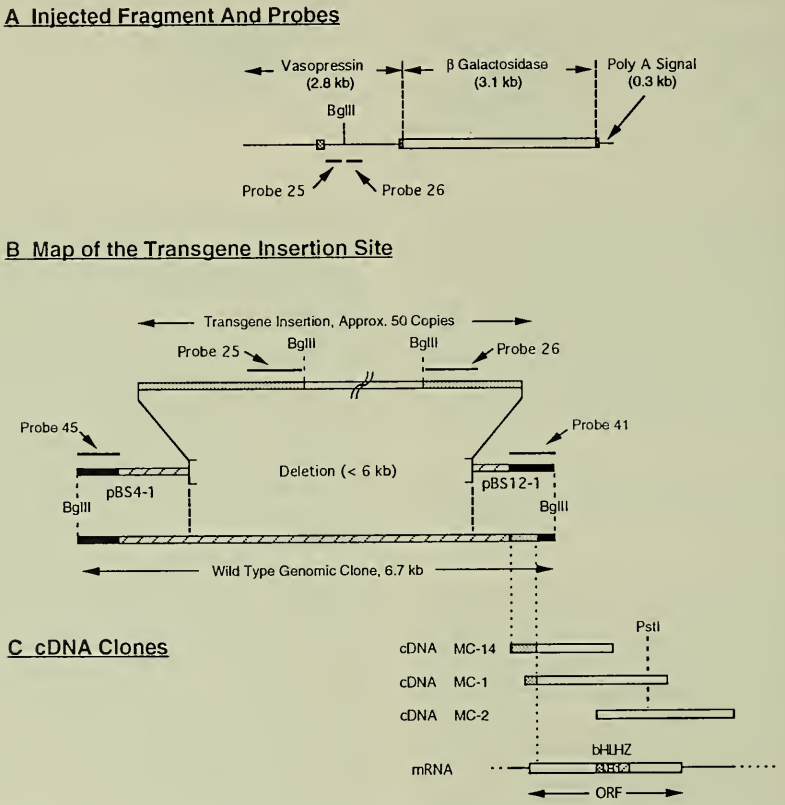


FIGURE 1. Transgene construct and organization of the transgene insertion site. (A) The mouse vasopressin/*Escherichia coli* β-galactosidase construct used for the generation of the VGA-9 transgenic line. Probes 25 and 26 abut the single *Bgl*II site and have been used to clone two flanking genomic sequences. (B) Schematic representation of the two flanking clones and a wild-type clone spanning the insertion site. The flanking clone pBS4-1 contains a unique genomic sequence from which probe 45 was derived, and the flanking clone pBS12-1 contains a unique genomic sequence from which probe 41 was derived. These probes both hybridize with a 6.7-kilobase *Bgl*II fragment of wild-type genomic DNA. The transgene insertion is accompanied by a deletion of <6 kilobases of genomic DNA. (C) Probe 41 was used to clone the cDNA MC-1 from a melanocyte library, and MC-1 was used to clone additional cDNAs that span the entire open reading frame of the Mi bHLH-Zip protein.

Genomic liver DNA from a VGA-9 homozygous mouse was digested to completion with *Bgl*II (an enzyme that cuts the injected fragment at a single site; Figure 1A) and the fragments were cloned into a lambda vector. The libraries were screened with the two probes, 25 and 26, that flank the *Bgl*II site (Figure 1A). Clones that were positive with either of the two probes, but not both, were further analyzed. This strategy was based on the assumption that ~50 transgene fragments were integrated in tandem and double-positive clones would represent internal *Bgl*II fragments composed only of transgene sequences; the elimination of these double-positives would thus allow for a relative enrichment of clones containing genomic flanking sequences. The *Bgl*II fragments from six single-positive clones were subcloned into a plasmid and sequenced from both ends. Two subclones, pBS4-1 and pBS12-1, contained sequences corresponding to portions of the transgene on one end and sequences unrelated to the transgene on the other (data not shown). Two new probes, 41 and 45, were then derived from these transgene-unrelated sequences thought to contain genomic flanking DNA.

If probes 41 and 45 were indeed derived from genomic flanking sequences, they should both detect altered genomic restriction fragments in VGA-9. This was in fact the case, as determined by Southern blot analysis performed with probes 41 and 45 and DNAs from wild type (+/+) at *mi*, heterozygous (*tg*/+) and homozygous (*tg*/*tg*) for the transgene-induced mutation (data not shown).

Isolation of *mi* cDNAs and Primary Structure of the Mi Protein

Southern analysis performed with DNA from other species besides mouse suggested that probe 41 contains exonic sequences. In fact, in Northern blots of total RNA from a mouse melanocyte cell line, melan-c, probe 41 detected a single band of approximately 5.5 kilobases (data not shown). Therefore, probe 41 was used to screen a melan-c cDNA library. A first clone, MC-1 (Figure 1C), was obtained and used to rescreen the library, which yielded two additional clones, MC-2 and MC-14. These clones were subsequently shown to map to the *mi* locus on chromosome 6 by interspecific backcross analysis (data not shown).

The three overlapping cDNAs spanned 1922 base pairs of sequence and contained a single large, 1257-base pair open reading frame beginning with a methionine codon at nucleotide positions 130 to 132 and ending with a cysteine codon at positions 1384 to 1386 (Figure 2). Homology searches indicated that this putative 413-amino acid Mi protein is a novel member of the basic helix-loop-helix–leucine zipper (bHLH-Zip) family of transcription factors. The locations of the characteristic basic DNA-binding domain and the HLH and leucine zipper domains of the Mi protein are highlighted in Figure 2. The Mi protein sequence also contains a number of potential phosphorylation sites, including a cyclic nucleotide-dependent protein kinase phosphorylation site at amino acid 409, protein kinase C phosphorylation sites at residues 404 and 405, and multiple casein kinase II phosphorylation sites.

The MC-14 cDNA contained an 18-nucleotide insertion (nucleotides 684 to 701, marked by lowercase letters in Figure 2) relative to MC-1 that introduced 6 amino acids into the putative Mi coding region. The 6-amino acid insertion is located just upstream of the Mi basic region. Reverse transcriptase-polymerase chain reaction (RT-PCR) amplification of RNA extracts from several wild-type mouse tissues suggests that this insertion resulted from alternative splicing. Another member of the bHLH-Zip protein family, Max, also encodes an alternatively spliced product containing a 9-amino acid insertion adjacent to the N-terminus of the basic region. The role of this alternatively spliced product in Max function has not yet been determined.

Sequence Comparison with Other bHLH-Zip Proteins

Among the many bHLH-Zip proteins, the predicted Mi sequence shows the highest degree of homology to TFE3 and TFEB (Figure 3; data not shown). TFE3 was first identified in humans as a protein that binds the E3 box of the μ-chain immunoglobulin enhancer. Its mouse homolog was found to bind additional transcriptional control elements in lymphoid-specific, muscle-specific, and ubiquitously expressed genes. TFEB was cloned from a human B-cell line expression library with a probe containing the binding site for the adenovirus major late transcription factor. Both TFE3 and TFEB are thought to be ubiquitously expressed. In in vitro assays, they both bind the same 6-base pair palindromic sequence CACGTG either as homodimers or as TFE3/TFEB heterodimers.

```
   1 TCG GGA TGC CTT GTT TAT GGT GCC TTC TTT ATG CCG GTC GCT CTT CGA ATT GGA ATT ACA GAA AGT AGA GGG AGG AGG ACT AAG TGG TCT   90

  91 GCG GTG TCT CCT GGG CTT GGG GCT GCC TGA AAC CTT GCT ATG CTG GAA ATG CTA GAA TAC AGT CAC TAC CAG GTG CAG ACC CAC CTG GAA  180
   1                                              Met Leu Glu Met Leu Glu Tyr Ser His Tyr Gln Val Gln Thr His Leu Glu   17

 181 AAC CCC ACC AAG TAC CAC ATA CAG CAA GCT CAG AGG CAC CAG GTA AAG CAG TAC CTT TCT ACC ACT TTA GCA AAT AAA CAT GCC AGC CAA  270
  18 Asn Pro Thr Lys Tyr His Ile Gln Gln Ala Gln Arg His Gln Val Lys Gln Tyr Leu Ser Thr Thr Leu Ala Asn Lys His Ala Ser Gln   47

 271 GTC CTG AGC TCA CCA TGT CCA AAC CAG CCT GGC GAC CAT GCC ATG CCA CCA GTG CCG GGG AGC AGC GCA CCC AAC AGC CCT ATG GCT ATG  360
  48 Val Leu Ser Ser Pro Cys Pro Asn Gln Pro Gly Asp His Ala Met Pro Pro Val Pro Gly Ser Ser Ala Pro Asn Ser Pro Met Ala Met   77

 361 CTC ACT CTT AAC TCC AAC TGT GAA AAA GAG GCA TTT TAT AAG TTT GAG GAG CAG AGC AGG GCA GAG AGT GAG TGC CCA GGT ATG AAC ACG  450
  78 Leu Thr Leu Asn Ser Asn Cys Glu Lys Glu Ala Phe Tyr Lys Phe Glu Glu Gln Ser Arg Ala Glu Ser Glu Cys Pro Gly Met Asn Thr  107

 451 CAC TCT CGA GCG TCG TGC ATG CAG ATG GAT GAT GTA ATT GAT GAC ATC ATC AGC CTG GAA TCA AGT TAT AAT GAA GAA ATT TTG GGC TTG  540
 108 His Ser Arg Ala Ser Cys Met Gln Met Asp Asp Val Ile Asp Asp Ile Ile Ser Leu Glu Ser Ser Tyr Asn Glu Glu Ile Leu Gly Leu  137

 541 ATG GAT CCG GCC TTG CAA ATG GCA AAT ACG TTA CCC GTC TCT GGA AAC TTG ATC GAC CTC TAC AGC AAC CAG GGC CTG CCA CCG CCA GGC  630
 138 Met Asp Pro Ala Leu Gln Met Ala Asn Thr Leu Pro Val Ser Gly Asn Leu Ile Asp Leu Tyr Ser Asn Gln Gly Leu Pro Pro Pro Gly  167

 631 CTT ACC ATC AGC AAC TCC TGT CCA GCC AAC CTT CCC AAC ATA AAA AGG GAG CTc aca gcg tgt att ttc ccC ACA GAG TCT GAA GCA AGA  720
 168 Leu Thr Ile Ser Asn Ser Cys Pro Ala Asn Leu Pro Asn Ile Lys Arg Glu Leu Thr Ala Cys Ile Phe Pro Thr Glu Ser Glu Ala Arg  197

 721 GCA TTG GCT AAA GAG AGG CAG AAA AAG GAC AAT CAC AAC TTG ATT GAA CGA AGA AGA AGA TTT AAC ATA AAC GAC CGC ATT AAG GAG CTA  810
 198 Ala Leu Ala Lys Glu Arg Gln Lys Lys Asp Asn His Asn Leu Ile Glu Arg Arg Arg Arg Phe Asn Ile Asn Asp Arg Ile Lys Glu Leu  227

 811 GGT ACT CTG ATC CCC AAG TCA AAT GAT CCA GAC ATG CGG TGG AAC AAG GGA ACC ATT CTC AAG GCC TCT GTG GAC TAC ATC CGG AAG TTG  900
 228 Gly Thr Leu Ile Pro Lys Ser Asn Asp Pro Asp Met Arg Trp Asn Lys Gly Thr Ile Leu Lys Ala Ser Val Asp Tyr Ile Arg Lys Leu  257

 901 CAA CGG GAA CAG CAA CGA GCT AAG GAC CTT GAA AAC CGA CAG AAG AAG CTG GAG CAT GCG AAC CGG CAC CTG CTG CTC AGA GTA CAG GAG  990
 258 Gln Arg Glu Gln Gln Arg Ala Lys Asp Leu Glu Asn Arg Gln Lys Lys Leu Glu His Ala Asn Arg His Leu Leu Leu Arg Val Gln Glu  287

 991 CTG GAG ATG CAG GCT AGA GCG CAT GGA CTT TCC CTT ATC CCA TCC ACC GGT CTC TGC TCG CCT GAT CTG GTG AAT CGG ATC ATC AAG CAA 1080
 288 Leu Glu Met Gln Ala Arg Ala His Gly Leu Ser Leu Ile Pro Ser Thr Gly Leu Cys Ser Pro Asp Leu Val Asn Arg Ile Ile Lys Gln  317

1081 GAA CCA GTT CTT GAG AAC TGC AGC CAG GAA CTT GTA CAG CAC CAG GCA GAC CTG ACA TGT ACG ACA ACT CTG GAT CTC ACG GAC GGT ACC 1170
 318 Glu Pro Val Leu Glu Asn Cys Ser Gln Glu Leu Val Gln His Gln Ala Asp Leu Thr Cys Thr Thr Thr Leu Asp Leu Thr Asp Gly Thr  347

1171 ATC ACC TTT ACC AAC AAC CTC GGC ACC ATG CCG GAG AGC AGC CCG GCC TAC AGC ATC CCC AGG AAG ATG GGC TCC AAC TTG GAA GAC ATC 1260
 348 Ile Thr Phe Thr Asn Asn Leu Gly Thr Met Pro Glu Ser Ser Pro Ala Tyr Ser Ile Pro Arg Lys Met Gly Ser Asn Leu Glu Asp Ile  377

1261 CTG ATG GAC GAT GCC CTC TCA CCT GTT GGA GTC ACC GAC CCA CTG CTG TCA TCA GTG TCG CCA GGA GCT TCA AAA ACA AGC AGC CGG AGG 1350
 378 Leu Met Asp Asp Ala Leu Ser Pro Val Gly Val Thr Asp Pro Leu Leu Ser Ser Val Ser Pro Gly Ala Ser Lys Thr Ser Ser Arg Arg  407

1351 AGC AGT ATG AGC GCA GAA GAA ACG GAG CAT GCG TGT TAG CGA GCC TGC CTT GCT CTG CCT CTG CAC GGA CCG CTC CCT CTC TTC TTC AGG 1440
 408 Ser Ser Met Ser Ala Glu Glu Thr Glu His Ala Cys                                                                  419

1441 AGA CTT TAT AAT TTA CCT GAA GAG GTT TTC TTG ATA ATT TTC CTT TAA TAT GAA ATT TTC GTG CTT TAT CAG TAG CCC TGC ATA TAT TTT 1530

1531 ATT TTT AGA ATT TTG TGA GCC AGA CTT GTA TAT TCT ATT TTA CAA CTA CAA AGG CCT CCT AAG TAT TGT ACC TTC AGC GTG CAG TAT CTG 1620

1621 TGA ACT GAT TTC TCC AAG TGT GAG CTT TCT GAG CAA GGG GAT TTT TTT GCT TCA GAG AAG TAA GTG TCT GTC CGT TTT CAT TCA GGG GAA 1710

1711 AAC TTG GTG TGA GCT TTG TCT GTC TGT GAC CTC CTT TGA AAT TAA CAT GTA AAG TTT AAT TAC ACG AAT GTA AAG CAA CCA AAA GAA GAA 1800

1801 AAC AAA GAA GAT GAT GAT TAT GAT AGA AGA AGG AGA AGA GGA GGA AGA GGA AGA GGA GGA GGA AGA GGA GGA AGA AAG ATG GTG GGA AAG 1890

1891 ACA TCC AAG ACT TTC TCG CTT TCT AAT ACG CG                                                                        1922
```

FIGURE 2. Partial sequence of the *mi* mRNA and predicted sequence of the Mi protein. The *mi* mRNA contains an alternatively spliced exon marked by lowercase letters (positions 684 to 701) that is present in approximately half of RT-PCR–amplified products prepared from RNA from various wild-type tissues. The shaded box marks the 19-residue region rich in basic amino acids. The open boxes identify the two helical domains that are interrupted by the underlined loop region. The underlined leucines, spaced 7 amino acids apart, conform to the prediction of a leucine zipper domain.

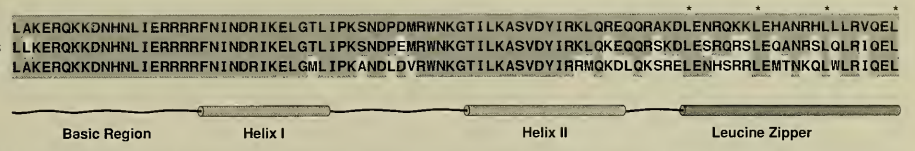


FIGURE 3. Comparison of the bHLH-Zip domains conserved between mouse Mi (mMi), mouse TFE3 (mTFE3), and human TFEB (hTFEB). The comparison shows near-perfect homology in the basic region and the helix I and helix II domains of the three proteins and the conservation of the leucines in the leucine zipper domain. Helix II and the leucine zipper may form a continuous α-helix as in Max.

Mi, mouse TFE3, and human TFEB have nearly identical basic regions and very similar HLH and Zip domains (Figure 3). This fact raises the interesting possibility that Mi may be able to form stable heterodimers with TFE3 and TFEB. The sequence similarity between these three proteins is not restricted to the bHLH-Zip domain; Mi and TFE3 share 61% identity (81% similarity) in the region between the N terminus and the basic domain, and 38% identity (55% similarity) between the zipper domain and the C terminus. With TFEB, these numbers are 44% identity (61% similarity) and 26% identity (44% similarity), respectively.

Expression of *mi* mRNA in the Adult

The pattern of *mi* mRNA expression in adult mouse tissues was determined by probing Northern blots of polyA$^+$ RNA isolated from various adult wild-type and VGA-9 (*tg/tg*) tissues with the *mi* cDNA probe MC-1 (not shown). Messenger RNA of the mouse mast cell line MC-6 and total RNA from melan-c cells served as controls. Two transcripts, 5.5 and 5.7 kilobases in size, were detected. The 5.5-kilobase transcript was exclusively expressed in melan-c cells and skin. The 5.7-kilobase transcript was expressed at high levels in heart, less prominently in skeletal muscle, and at barely detectable levels in various other organs. It was also expressed in the mast cell line MC-6, raising the possibility that the low levels of *mi* expression in various tissues could reflect the presence of mast cells. Identical expression patterns were evident with a probe derived from the *mi* 3′ untranslated region, suggesting that none of these transcripts represents cross-hybridization to transcripts originating from a related gene (data not shown). *mi* expression was not detected in VGA-9 (*tg/tg*) mice, indicating that transgene insertion had disrupted expression of the *mi* gene.

In contrast to what was observed with MC-1 and probes corresponding to the 3′ untranslated region of *mi*, probe 41, which contains both 5′ untranslated sequences and the N terminus of the coding region, only detected the 5.5-kilobase transcript seen in skin and melan-c RNA. These findings suggest that the two differently sized *mi* mRNA transcripts may result in part from alternative splicing at the 5′ end of the *mi* mRNA and that *mi* expression seen in skin may reflect the presence of coat melanocytes.

Expression of *mi* in Development

If we have indeed cloned the *mi* gene, we would expect this gene to be expressed at the sites and developmental stages first affected by *mi*. It has been shown previously that *mi* mice lack melanocytes in the skin and in hair follicles. It has also been shown that in both *mi/mi* and Mi^{wh}/Mi^{wh} embryos, the outer layer of the optic cup (normally the future pigment layer of the retina) thickens abnormally at embryonic day (E) 10 to 11, leading to failure of the choroid fissure to close and to colobomatous microphthalmia. In addition, it has been shown that VGA-9 mice have an abnormal inner ear with a stria vascularis devoid of melanocytes. These melanocytes are thought to originate in the neural crest, from which they migrate to areas surrounding the otic vesicle (the future inner ear) before they become concentrated in the stria vascularis. For these reasons, we concentrated our in situ hybridization analysis of *mi* expression on the developing eye, ear, and skin.

Using an MC-1 antisense riboprobe, we observed prominent expression of *mi* RNA in wild-type albino embryos (possessing melanocytes, but lacking the melanin granules that might interfere in these assays) in the pigment

layer of the retina (Figure 4a, d), individual cells surrounding the otic vesicle (Figure 4g), and hair follicles (Figure 4k). We found a similar distribution of signal in serial sections hybridized with an antisense riboprobe corresponding to the 3′ untranslated region of *mi* as well as with an antisense riboprobe specific for TRP-2/DT, a melanoblast marker (data not shown). This latter result and the restricted location and small number of MC-1 antisense riboprobe-positive cells suggested that they are melanoblasts. Sense riboprobes gave no signal (Figure 4b, e, h, l). In VGA-9 transgenic embryos, *mi* was not expressed in any tissue (Figure 4i, m; and data not shown) except the pigment layer of the retina (Figure 4c, f). Here, we detected a much reduced, though specific, signal that was also evident with an antisense riboprobe for the 3′ untranslated region (data not shown). As shown above, the transgene insertion site in VGA-9 mice lies upstream of the known coding region of *mi* (Figure 1). As we have not yet fully characterized the 5′ end of the *mi* gene and the 5′ end of the *mi* mRNA appears to undergo alternative splicing, we do not know whether the transgene insertion is located upstream of the *mi* promoter, within the *mi* 5′ untranslated region, or within an alternatively used 5′ coding region. It is thus possible that despite transgene insertion, some residual *mi* expression still occurs, although it is apparently restricted to the retina. Alternatively, this retinal expression could be due to a cryptic promoter activity of the adjacent transgene sequences.

In addition to expression in presumptive melanoblasts, we also observed *mi* expression in embryonic heart (data not shown), consistent with the observation that adult heart also expresses *mi*. Heart abnormalities have not been reported for *mi* mice, although this obviously requires further evaluation.

Molecular Analysis of Two Independent *mi* Alleles

Further evidence that we have cloned the *mi* gene comes from analysis of two additional *mi* alleles, mi^{ws} and *mi*. Five overlapping cDNA probes that collectively span the *mi* coding region were generated by PCR amplification of the *mi* cDNAs (Figure 5A). These probes, in addition to probe 41, were hybridized to *Bgl*II-digested mouse DNA in order to construct a limited wild-type genomic restriction map of *mi*-hybridizing fragments (Figure 5B). Each probe was hybridized to a Southern blot of wild-type and homozygous mi^{ws} genomic DNA (Figure 5C). Probe 41 hybridized to a 6.7-kilobase *Bgl*II fragment in both mi^{ws} and wild-type DNA. In contrast, probe 1, which spans nucleotides 214 to 563, recognized 9.6-, 4.8-, and 1.6-kilobase fragments in wild-type DNA, while only a 9.6-kilobase fragment was detected in mi^{ws} DNA. Similarly, probe 2, which spans nucleotides 478 to 860, detected 9.6-, 4.8-, and 3.9-kilobase fragments in wild-type DNA, while only 9.6- and 3.9-kilobase fragments were detected in mi^{ws} DNA. Probes 3, 4, and 5 recognized identically sized fragments in both mutant and wild-type DNA (data not shown). These data strongly suggest that the mi^{ws} mutation results from a discrete intragenic deletion in the N-terminal coding region of the *mi* gene.

Southern analysis of genomic DNA of mice carrying the original *mi* allele failed to identify any *mi*-associated alterations. To determine whether the coding region of this allele contained a small structural alteration that might not be detected by Southern analysis, we amplified the *mi* coding region between nucleotides 234 to 1772 by PCR using reverse-transcribed heart and kidney RNA. Subsequent sequence analysis identified four differences between the original *mi* and the wild-type *mi* sequences shown in Figure 2. All differences were confirmed in multiple independent experiments, indicating that they are not PCR artifacts. Two differences were found in the 3′ untranslated region of the gene and represent a polymorphism also observed in other wild-type mouse strains (data not shown). Two additional differences were found within the coding region. One of them was an A to G transition at position 1137, which is silent, and the other was a 3-nucleotide deletion (positions 214 to 217 in Figure 2) that results in the loss of an arginine residue from a series of four conserved, consecutive arginines (Figure 6). These arginine residues are located at the C terminus of the *mi* basic region (Figures 2 and 3), which, based on mutational studies of TFEB and the crystallographic data of Max, might be involved in contacting DNA. The phasing of the basic region with respect to helix 1 is altered by this deletion, which is likely to have a significant effect on the ability of the mutant protein to bind DNA. Thus, two independent alleles at *mi* are expected to encode proteins with severely impaired function.

Collectively, these results demonstrate that we have identified a gene that encodes a novel bHLH-Zip protein

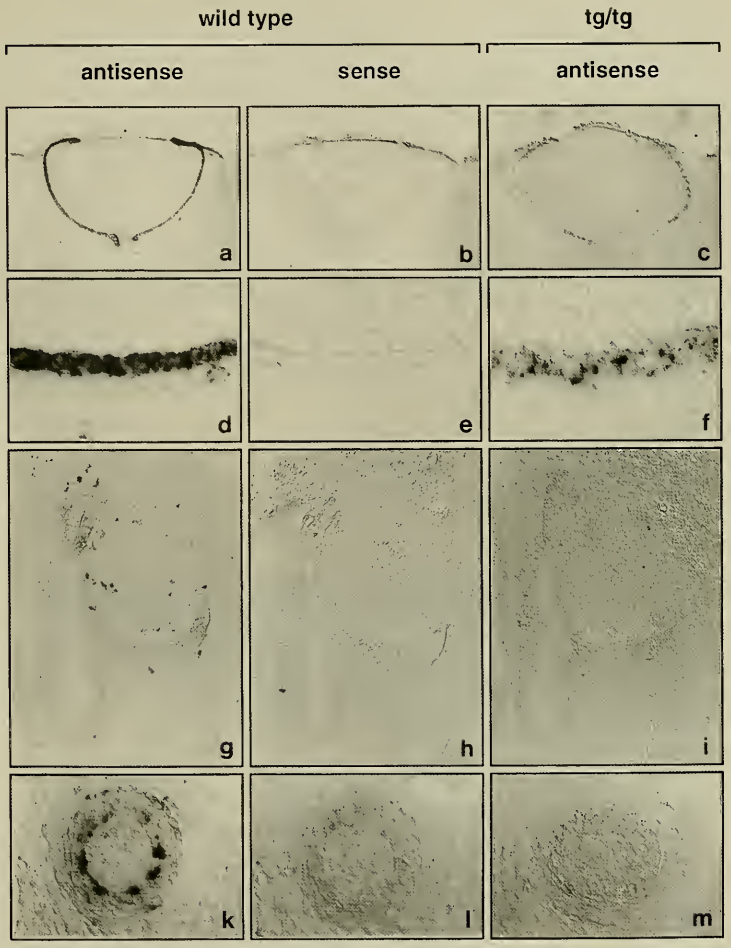


FIGURE 4. Expression of *mi* RNA in wild-type and VGA-9 (*tg/tg*) embryos. Cryostat sections of wild-type albino (a, d, g, k, and corresponding adjacent sections b, e, h, l) and VGA-9 (c, f, i, m) embryos harvested on day 13.5 (a-i) and 16.5 (k-m) of gestation. The sections were hybridized either with an MC-1 antisense or MC-1 sense riboprobe as indicated. (a) *mi* RNA is expressed in the outer layer of the retina of a wild-type eye when probed with the antisense probe. (b) No signal is evident on the adjacent section probed with the sense probe. (c) Weak signal is evident in the outer retinal layer of a VGA-9 eye. (d-f) Higher magnifications of the retina sections shown in panels a-c. (d) Note cuboidal single cell layer expressing *mi* RNA in wild-type outer retinal layer. (e) Sense riboprobe control. (f) In the VGA-9 retina, oblongated cells of the outer retinal layer show weak staining. (g) Presumptive melanoblasts in the vicinity of the otic vesicle are stained. (h) Sense riboprobe control. (i) Absence of staining in VGA-9 section, suggesting absence of presumptive melanoblasts. (k) Presumptive melanoblasts in wild-type hair follicle (whisker). (l) Sense riboprobe control. (m) VGA-9 hair follicle probed with antisense riboprobe. Magnifications: (a-c) 53X; (d-f) 350X; (g-i) 78X; (k-m) 225X.

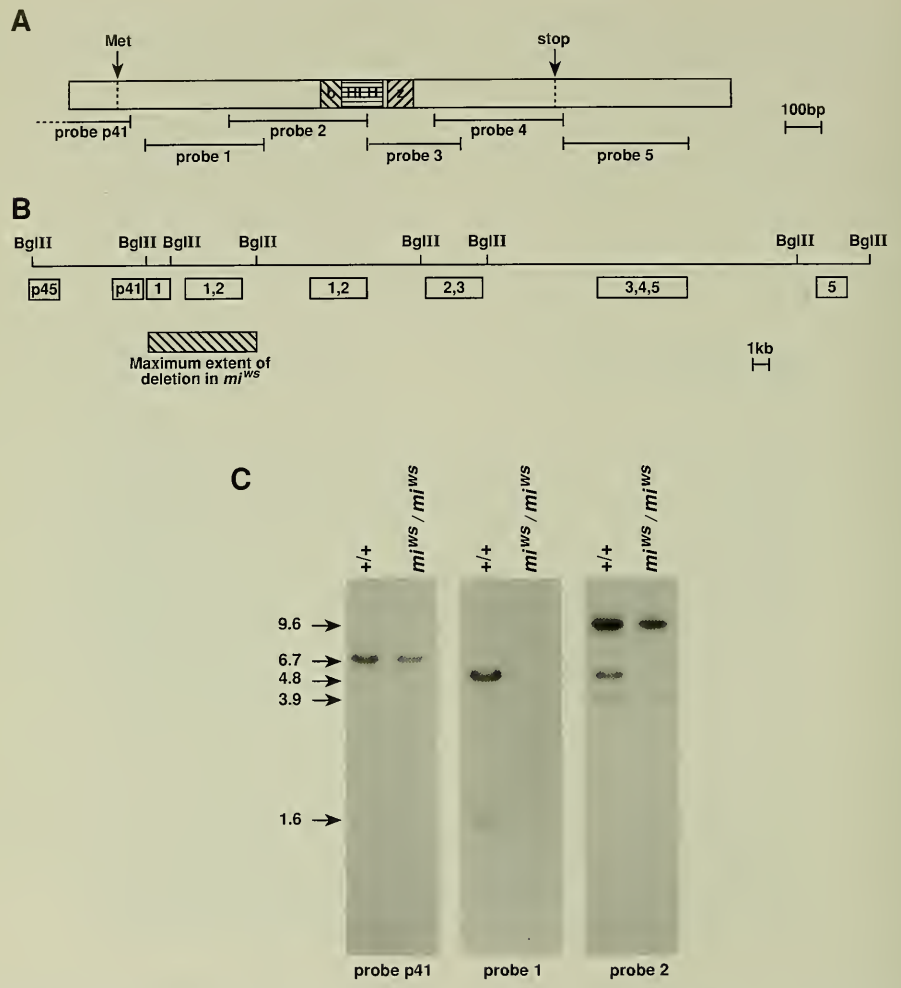


FIGURE 5. Genomic analysis of the *mi*^{ws} allele. (A) Schematic representation of the *mi* cDNA sequence, showing the relative positions of putative initiator methionine and TAG stop codons as well as basic (b), helix-loop-helix (HLH), and leucine zipper (z) domains. The locations of five PCR-generated probes are indicated by bars below. The origin of probe 41 is described in the text. The probes span the following nucleotides: probe 1, 214 to 563; probe 2, 460 to 860; probe 3, 873 to 1133; probe 4, 1117 to 1432; probe 5, 1412 to 1790. (B) An exonic restriction map of *Bgl*II-digested C57BL/6J DNA. Probes that hybridize to a given *Bgl*II genomic restriction fragment are indicated in a box below each fragment. The maximum extent of the *mi*^{ws} deletion is indicated by the hatched box. (C) Comparative Southern analysis of C57BL/6J +/+ and C57BL/6J-*mi*^{ws}/*mi*^{ws} DNA hybridized with probes 41, 1, and 2. The size of each hybridizing fragment (in kilobases) is indicated on the left.

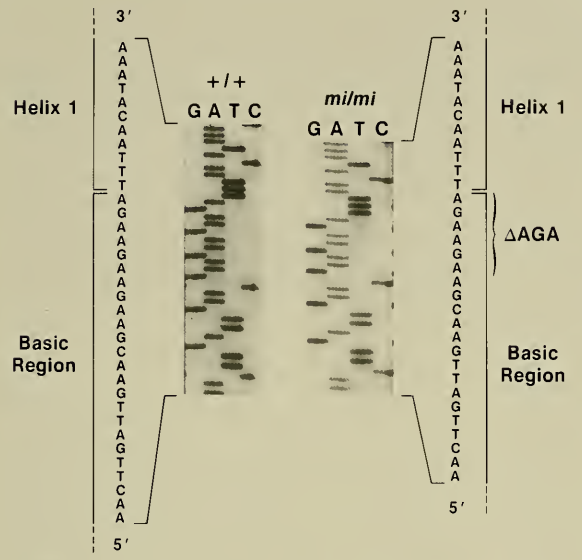


FIGURE 6. The *mi* allele contains a 3-nucleotide deletion in the *mi* basic region. The sequence of the 3′ end of the *mi* basic region along with the beginning of helix 1 from the wild-type allele is shown in the left panel. The right panel depicts the corresponding sequence from *mi* mice. Note the three trinucleotide AGA repeats encoding arginine residues in the wild-type sequence. Only two of the three AGA repeats are present in the *mi* sequence. The position of this 3-nucleotide deletion is indicated on the far right.

and provide strong evidence that this gene is, in fact, mouse *mi*. First, in two independent *mi* alleles, the gene is mutated in its coding region. Second, in a line of mice with a typical *mi* phenotype due to an insertional mutation at *mi*, expression of this gene is virtually abolished. Third, the gene is expressed in the developing eye, inner ear, and skin precisely at locations and in cell types affected by *mi*. Fourth, the gene encodes a putative transcription factor, which is consistent with the notion that the *mi* gene product acts in a cell-autonomous fashion.

The Molecular Nature of the *mi* and mi^{ws} Mutations

Southern analysis of genomic DNA indicates that the spontaneous mi^{ws} mutation resulted from an intragenic deletion within the 5′ coding region of the *mi* gene. The fact that the mi^{ws} allele is a strong semidominant allele; that some mi^{ws}/mi^{ws} homozygotes, although white, are normophthalmic; and that Mi^{wh}/mi^{ws} compound heterozygotes have more pigmentation and more normal eyes than homozygotes of either allele suggest that mi^{ws} expresses a protein with some residual activity. Future studies including the use of specific antibodies will help to clarify this point.

The original *mi* allele is also semidominant but, compared to mi^{ws}, phenotypically more severe. Molecularly, the mutation in the *mi* allele is characterized by the deletion of a single arginine in a row of four conserved arginines at the C terminus of the DNA-binding basic domain. The structural and mutational analysis of the close relative TFEB and the crystal structure of the related Max protein indicate that the basic region of these proteins assumes the conformation of an α-helix whose C terminus contacts the major groove of DNA through a face of the helix composed of critically conserved residues. In Max, all conserved residues in this region make either base or phosphate contact. In the basic region of the Mi mutant protein, the deletion of an arginine alters the phasing of the α-helix in this region, hence changing some of the amino acids on the face of the helix critical for binding DNA. Thus, *mi* may encode a protein incapable of binding DNA, although still capable of forming dimers.

Is *mi* a Complex Locus?

The variability in phenotypes produced by different microphthalmia alleles and the phenomenon of interallelic complementation have led to the notion that the *mi* locus is complex, i.e., the complete microphthalmia phenotype might result from defects in more than one linked gene. However, our results with *mi* seem to rule out this possibility. As discussed above, the *mi* mutation is due to an in-frame deletion of a single arginine residue in the Mi basic region. Yet, *mi* mice display a combination of all phenotypic traits that are associated with mutations at this locus.

The simplest explanation of this observation is that the various *mi* phenotypes are produced by mutations in a single gene and that the phenotypic variabilities simply reflect allelic differences.

Protein Dimerization and Interallelic Complementation

HLH proteins are well known for their propensity to form specific homo- and heterodimers and to depend on dimerization for binding DNA. Myc, for example, is an unstable protein that as a homodimer binds DNA poorly; its activity crucially depends on heterodimerization with Max and is regulated by formation of additional heterodimers with the Mad and Mxi-1 proteins. The close relationship of Mi with TFE3 and TFEB, two bHLH-Zip proteins that form homodimers and TFE3/TFEB heterodimers, suggests that Mi, too, will form homodimers and possibly heterodimers with these proteins.

Protein dimerization provides a rational basis for explaining the phenomenon of interallelic complementation and enhancement. For such interallelic interactions to occur, however, neither of the two mutant proteins may be entirely nonfunctional. As discussed earlier, the *mi* allele is expected to encode a protein with a severely reduced capacity to bind DNA, and yet complementation is observed with this allele. In the absence of DNA binding, a simple model involving only direct interactions between mutant Mi proteins could not easily explain the complementation with *mi*. Therefore, either the mutant protein encoded by the *mi* allele must retain some capacity to bind DNA or Mi is part of a complex network of interacting transcription factors. In any event, our finding that a genetic locus known for its interallelic interactions encodes a bHLH-Zip protein suggests that dimerization between these transcription factors operates not only in vitro but also in vivo.

Previous work has suggested that the cellular defects associated with mutations at *mi*, in particular those of melanocytes, result from reduced cell proliferation, disturbed migration, blocked differentiation, or premature cell death. Therefore, the Mi protein may affect one or several of these processes. Interestingly, the bHLH-Zip protein Myc is a transcription factor that, when dysregulated, is oncogenic and can block differentiation and induce apoptosis. Thus, Mi protein and other members of this protein family may share not only important structural domains but perhaps also some functional properties.

It is still far from clear which target genes are being regulated by any of the mammalian bHLH-Zip proteins, including Mi. It has been reported that *mi* mast cells display low numbers of c-Kit receptors, which suggests that Mi regulates the expression of *c-kit*. However, based on results obtained in an in vitro mast cell/fibroblast coculture system, it has been proposed that *mi* represents a common "downstream" target of the *Sl*/Kit and CSF-1/*c-fms* signal transduction pathway. Now that *mi* has been cloned, direct experiments to address these issues can be performed.

Secondary Bone Resorption and *mi*

Deficiencies in secondary bone resorption (osteopetrosis) are observed with three microphthalmia alleles, *mi*, Mi^{or}, and mi^{di} (defective iris). The earlier analysis of the osteopetrotic defect combined with the results presented here suggests that, in these mice, osteopetrosis is due to defects in the *mi* gene. However, it is unclear whether these alleles are amorphic mutations, i.e., encode totally nonfunctional proteins, or whether they are neomorphic mutations, i.e., encode proteins still capable of forming heterodimers with other proteins and, hence, capable of acting in a dominant-negative fashion or binding to additional DNA motifs not normally recognized by Mi.

Indirect evidence implicating bHLH-Zip proteins in secondary bone resorption has been reported. Mice in which the gene for Fos, a basic region-leucine zipper (bZip) protein, has been rendered nonfunctional by homologous recombination develop osteopetrosis with deficiencies in bone remodeling and tooth eruption reminiscent of the osteopetrotic phenotype of *mi* mutations. A bHLH-Zip protein designated FIP (Fos-interacting protein) was recently identified by its ability to form heterodimers with Fos. Thus, it is conceivable that Mi also forms heterodimers with either Fos or FIP and that the osteopetrotic phenotype in *mi* is mediated through these proteins.

Mouse *mi* and Human Disease

Microphthalmia, pigment alterations with or without associated hearing impairment, and bone abnormalities are symptoms found in many human hereditary disorders.

An example of a disorder characterized by pigment alterations in skin and eye and variable congenital sensorineural deafness is the genetically heterogeneous Waardenburg syndrome (WS). Recently, *PAX-3*, on human chromosome 2, was found mutated in six independent WS families; however, linkage to a locus on chromosome 2 could not be established for many WS families. For these families, mutations in the human homolog of mouse *mi* remain candidates for the origin of the syndrome.

One of the *mi* alleles, mi^{vit}, has been proposed as a model for human vitiligo, a disorder characterized by progressive elimination of melanocytes in the skin. Homozygous mi^{vit}/mi^{vit} mice show a progressive dilution of the pigmented areas with each molt until, in old age, they are completely white. As in vitiligo, this progressive loss of pigmentation correlates with a reduction in the number of functional melanocytes in hair follicles. As the mouse *mi* gene is evolutionarily well conserved, it should now be possible to clone the corresponding human gene and to determine whether this gene has any role in human disease.

Molecular Genetics of Development Section

Nancy A. Jenkins

Karen Avraham
Mary Bedell
Brian Cho
Bryn Eagleson
Colin Fletcher
Debra Gilbert
Carolyn Hustad
Karen Moore
William Perry
Eirikur Steingrimsson
Deborah Swing

The focus of the Molecular Genetics of Development Section is to identify and characterize genes important in normal mammalian development. Classic genetic analysis of variant mouse phenotypes has defined numerous genetic loci affecting development and, in some cases, numerous alleles at a given locus. However, the physica basis for these mutations and an understanding of the impact of mutagenic events on gene expression and on normal developmental processes have been impeded by difficulties in gaining molecular access to the mutated genes. The murine dilute (*d*)-short ear (*se*) region of mouse chromosome 9 is among those rare intervals for which extensive analysis at both the genetic and molecular levels is possible. This combined molecular and genetic analysis has elucidated the gene product of the *d* locus. This year, in collaboration with the Molecular Genetics of Oncogenesis Section, Liane Russell (Oak Ridge National Laboratory, Oak Ridge, TN), and David Kingsley and associates (Stanford University Medical School), we report the product of the *se* locus.

The Mouse *Short Ear* Skeletal Morphogenesis Locus Is Associated with Defects in a Bone Morphogenetic Member of the TGF-β Superfamily

Mice offer a unique opportunity to combine genetic and biochemical studies of skeletal growth and patterning. At least 150 mouse loci are known that affect various aspects of skeletal morphogenesis. One of these genes, the mouse *short ear* locus on chromosome 9, has a particularly rich history of both developmental and genetic study. Mice carrying a recessive mutation that causes a marked reduction in the size of the external ear were first reported over 70 years ago. Subsequent studies showed that the classical *short ear* mutation also causes a broad range of internal skeletal defects, including loss of several small bones, alterations in size or shape of the xiphoid process at the end of the sternum, reduction of ventral processes normally found specifically on the sixth cervical vertebra, deletion of one pair of ribs, and numerous other changes.

Developmental studies suggest that the morphological defects in short-eared animals arise prior to the overt differentiation of bone and cartilage. In normal animals, the earliest morphological sign of skeletal development is the aggregation of dispersed mesenchyme cells into regions of local high cell density (condensations). The cellular aggregates then synthesize large quantities of extracellular matrix molecules and ultimately differentiate into a cartilaginous or bony element surrounded by a fibrous sheath. Several of the morphological alterations in short-eared animals have been traced back to defects in the size, shape, and number of early condensations of mesenchyme. These observations suggest that the *short ear* gene is required for either the production or the interpretation of the early embryonic signals that cause mesenchyme cells to aggregate into the outline of future skeletal elements.

Genetic studies of the *short ear* locus have been greatly aided by the inclusion of this gene in the large specific-locus mutagenesis experiments begun in the 1940's. In these experiments, wild-type mice were treated with radiation or chemicals and bred to a special test strain homozygous for the *short ear* mutation and six distinct coat color markers.

As a result of the experiments, hundreds of new mutations have been isolated that affect the *short ear* gene, the closely linked *dilute* coat color marker, and surrounding chromosomal loci. Extensive complementation tests between these mutations has produced a detailed map of functional units within an interval of several centimorgans surrounding the *dilute* and *short ear* genes. Many of the radiation-induced mutations affect several different functional groups and are known to be deletion mutations.

We exploited the large set of mutant chromosomes available in the *short ear* region to direct a chromosome walk to the *short ear* locus. The mouse gene for bone morphogenetic protein-5 (*Bmp-5*) maps within this walk and is disrupted by several different *short ear* mutations. These results suggest that the various skeletal defects seen in short-eared animals are the results of defects in *Bmp-5*.

A Chromosome Walk in the *Short Ear* Region

We have previously used various deletions in the *dilute-short ear* region to show that a brain cDNA clone called *1B1075* maps in the very small 0.15-centimorgan (cM) interval between the *dilute* and *short ear* loci. Given the average relationship between genetic and physical distance in the mouse, the *1B1075* gene should thus be no more than 200 to 300 kilobases away from the *short ear* locus.

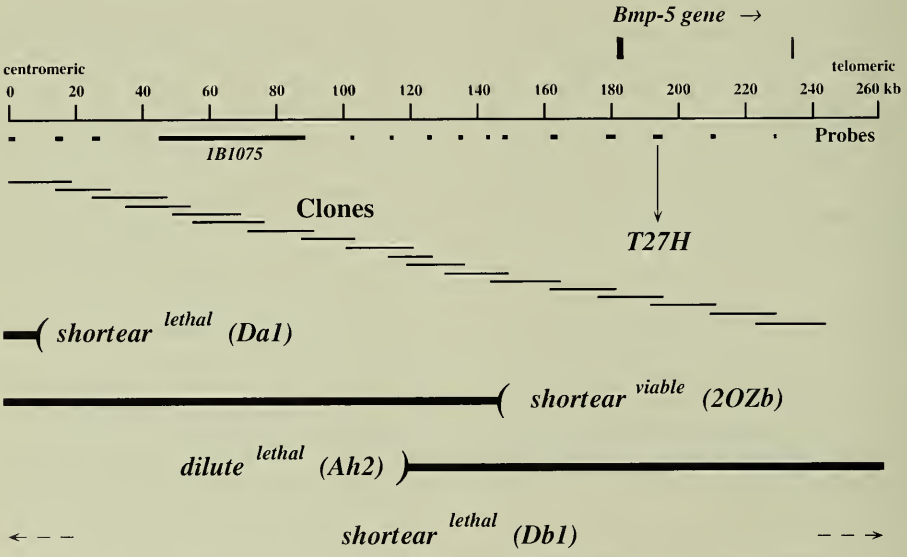


FIGURE 1. A 240-kilobase chromosome walk in the *short ear* region. A series of overlapping phage clones was isolated as described in the text. Unique-sequence probes were isolated at the indicated positions along the walk. The deletion breakpoints in the *Ah2*, *Da1*, and se^{2OZb} mutations are drawn between the two closest probes that did or did not hybridize to DNA samples from each mutation (see representative Southern blot studies in Figure 2). Thick bars and blank regions represent the DNA sequences that are present and deleted in each mutation, respectively. All of the probes analyzed appear to be missing in DNA samples from the *Db1* mutation. The unknown edges of the *Da1* deletion are thus indicated with dotted arrows at the bottom of the figure. The probe that recognizes an altered restriction fragment in the T27H translocation mutation is indicated with an arrow. The *Bmp-5* gene is shown above the chromosome walk. The black box corresponds to the 246-base pair *Bgl*II fragment that we have sequenced at the 5′ end of the *Bmp-5* gene (see Figure 7). The gray boxes represent genomic restriction fragments that hybridize to a probe containing the complete open reading frame of the human *BMP5* gene. Centromeric and telomeric orientation of the walk is provided by the breakpoints in the *Ah2* and *Da1* mutations, which have previously been ordered with respect to genetic maps of chromosome 9.

To reach the *short ear* gene, we used the *1B1075* probe to screen a phage genomic DNA library, made from the C57BL/6J strain, which is wild type at the *short ear* locus. Clones isolated in the initial screen were assembled into an overlapping set by restriction mapping. The total cloned region was then expanded in both directions by rescreening the genomic DNA library with probes that were isolated from the phage clones located at the two ends of the initial set (Figure 1).

At each subsequent walking step, new repeat-free probes were isolated and used to analyze DNA samples from wild-type mice and from mice carrying various mutations at or near the *short ear* locus. The starting point for the walk, the *1B1075* probe, was completely missing from three deletion mutations that are known to include multiple functional units on the complementation map and are lethal in the homozygous state (*Da1*, *Db1*, and *Ah2*; see Figure 2B). As the walk expanded, we identified the edge of deletions in both the *Da1* prenatal lethal mutation (which includes the *short ear* locus and functional units on the telomeric side of *short ear*) and the *Ah2* prenatal lethal deletion (which includes functional units on the centromeric side of the *short ear* locus but does not include the *short ear* gene itself) (Figures 1, 2A, and 2C). These breakpoints allowed us to orient the molecular walk with previous genetic maps of the *short ear* region. Subsequent walking steps were carried out only in the predicted direction of the *short ear* locus (telomeric in Figure 1).

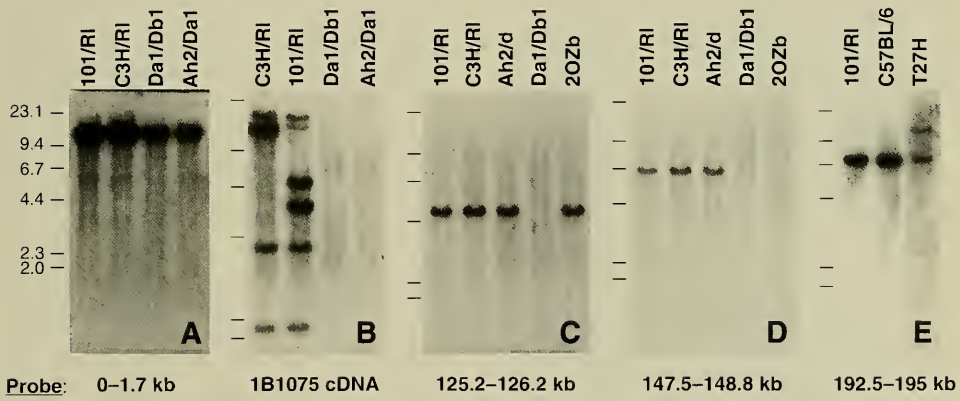


FIGURE 2. Southern blot analysis of mutations in the *short ear* region. Five micrograms of DNA from the indicated mouse strains was digested with various restriction enzymes and analyzed by Southern blot hybridization with chromosome-walking probes from Figure 1. The C3H/Rl, 10I/Rl, and C57BL/6 samples represent control strains. *Ah2/Da1*, *Da1/Db1*, and *Ah2/d* refer to compound heterozygotes between the *Da1*, *Db1*, *Ah2*, and *dilute* mutations. Although the large deletions in the *Da1*, *Db1*, and *Ah2* mutations are lethal when homozygous, they can be combined to generate viable long-eared (*Ah-2/Da1*) or short-eared (*Da1/Db1*) animals that are missing DNA sequences only in the region of overlap between the deletions. T27H DNA samples were isolated from long-eared mice heterozygous for the T27H translocation. The 20Zb samples were isolated from short-eared mice homozygous for the se^{20Zb} mutation. (A) A 1.7-kilobase *Bam*HI/*Sal*I fragment that begins at position 0 in Figure 1 recognizes a single *Eco*RI fragment in both *Da1/Db1* and *Ah2/Da1* DNA, suggesting that this probe is beyond the deletion breakpoint in *Da1*. (B) All *Bam*Hl fragments detected by the original *1B1075* cDNA probe are deleted in *Da1/Db1* and *Ah2/Da1* mice. These fragments are still present in mice carrying se^{20Zb}. (C) A 1 kilobase *Hin*dIII/*Sal*I fragment located between positions 125.2 and 126.2 in Figure 1 recognizes a single *Bgl*II fragment of wild-type size and intensity in *Ah2/d* and se^{20Zb} DNA, suggesting that this probe is telomeric to the deletion breakpoint in *Ah2* and centromeric to the deletion in se^{20Zb}. (D) A 1.3-kilobase *Kpn*I/*Sal*I fragment from positions 147.5 to 148.8 in Figure 1 detects a single *Sac*I fragment that is missing in DNA samples from the se^{20Zb} deletion, suggesting that this probe has crossed the deletion breakpoint in se^{20Zb}. (E) A 2.5-kilobase *Kpn*I/*Sal*I probe from positions 192.5 to 195 detects a novel *Bam*HI fragment in heterozygous carriers of the T27H mutation. Size markers (in kilobases) are indicated.

A Large Deletion in a Viable *Short Ear* Allele

The original *1B1075* probe did not detect any deletions or DNA rearrangements in a large collection of viable alleles at the *short ear* locus (9 radiation-induced and 17 chemical-induced mutations recovered in Oak Ridge mutagenesis studies, and two spontaneous mutations). Unlike the *Da1*, *Db1*, and *Ah2* deletions, these mutations produce only the *short ear* phenotype in the homozygous state. As the walk expanded in the telomeric direction, we eventually reached the edge of a deletion present in a viable *short ear* allele called se^{20Zb} (see position 145 in Figure 1). This mutation was originally isolated in a specific-locus mutagenesis experiment after treatment of female mice with 400 rads of acute X irradiation. The deletion present in se^{20Zb} was at least 84 kilobases long, as judged by Southern blot analysis of se^{20Zb} DNA samples with probes from additional chromosome walking steps (positions 145 to 229 in Figure 1).

In order to isolate the other edge of this deletion, we used a probe located at positions 142.7 to 143.4 in Figure 1 to isolate several deletion-breakpoint fusion clones from a partial-digest genomic DNA library constructed from mice carrying the se^{20Zb} deletion. DNA probes isolated from the telomeric edge of the resulting clones were then used to initiate a short walk on the other side of the se^{20Zb} deletion (see Figure 3). Both the original starting point for the walk (the *1B1075* probe) and a probe from the other end of the se^{20Zb} deletion (positions 709 to 709.5 in Figure 3) mapped to a large *Bss*HII fragment of approximately 725 kilobases in spleen DNA samples from multiple wild-type strains, including C57BL/6J, C3H/Rl, and 101/Rl (Figure 4A and data not shown). This fragment is reduced in size by approximately 550 kilobases in se^{20Zb} DNA, producing a 175-kilobase fragment recognized by both probes. One end of the 725-kilobase wild-type *Bss*HII fragment has been directly isolated in phage clones at position 27.5 kilobases in the genomic walk. An unmethylated

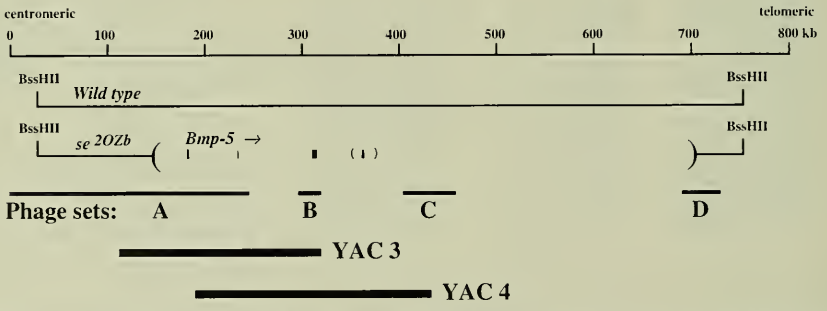


FIGURE 3. An expanded chromosome walk in the *short ear* region. The original 240-kilobase chromosome walk is shown as a single bar (phage set A). Two YAC clones were isolated that extend past the original walk. End probes from these YACs were used to isolate additional lambda walking clones (phage sets B and C). A probe located from positions 142.7 to 143.4 was used to isolate deletion breakpoint fusion clones from a se^{20Zb} genomic DNA library. These clones were used to isolate a small overlapping set of wild-type genomic clones located on the telomeric side of the se^{20Zb} mutation (phage set D). Phage sets B and C were positioned relative to phage set A using the known sizes and overlaps of the two YAC clones. Phage set D was positioned relative to phage set A by pulsed-field gel mapping. Probes from both sets map to the same 725-kilobase *Bss*HII fragment in wild-type DNA and the same 175-kilobase fragment in se^{20Zb} DNA (see Figure 4). A probe from the telomeric end of phage set D maps 25 kilobases away from an unmethylated *Bss*HII site in spleen DNA (detected by *Bss*HII cleavage of a 42-kilobase *Nde*I fragment recognized by this probe). A site that defines the centromeric end of the large *Bss*HII fragment has been isolated directly in phage DNA (position 27.5). The genomic alterations in the se^{20Zb} mutation can be explained by a 550-kilobase deletion located within the wild-type *Bss*HII fragment (see bars at top of figure). *Bmp-5* sequences map to several restriction fragments found near the centromeric end of the se^{20Zb} deletion (black and gray boxes as shown in Figure 1). The gray box shown in parentheses corresponds to a 2.5-kilobase *Eco*RI fragment recognized by a probe from the extreme C terminus of the *BMP5* coding region (nucleotides 1875 to 2062). This *Eco*RI fragment is found in DNA from YAC4, but not DNA from YAC3 or phage set B or C. Additional portions of the *Bmp-5* gene may map in the regions between phage sets A-B and B-C.

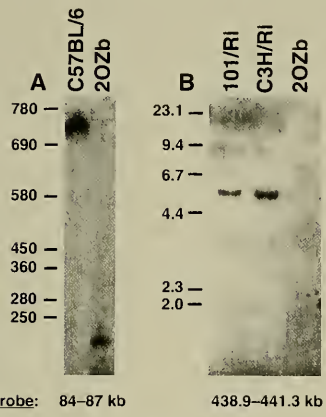


FIGURE 4. A large deletion in the se^{20Zb} mutation. Spleen DNA samples from C57BL/6J and se^{20Zb} homozygotes were digested with *Bss*HII, separated by pulsed-field gel electrophoresis, and analyzed by blot hybridization with a 3-kilobase genomic R1 fragment located between positions 84 and 87 in Figures 1 and 3 (panel A). The *Bss*HII fragment seen in the mutant is approximately 550 kilobases smaller than that seen in the control strain. A 1.4-kilobase genomic probe located between positions 438.9 and 441.3 in Figure 3 is missing in DNA samples from se^{20Zb}, confirming a large deletion in this mutation (panel B, *Spe*I digest).

*Bss*HII site has also been located approximately 25 kilobases away from the end of the small chromosome walk on the telomeric side of the se^{20Zb} deletion (see Figure 3 legend). The size of the altered *Bss*HII fragment and the position of these two known *Bss*HII sites relative to the se^{20Zb} breakpoints are both consistent with the presence of a 550-kilobase deletion in the se^{20Zb} mutant chromosome.

To isolate more of the DNA within this deletion, we used polymerase chain reaction (PCR) primers near position 195 in Figure 1 to isolate two yeast artificial chromosome (YAC) clones from a mouse genomic library also prepared from the C57BL/6J strain. These YACs were aligned with the existing genomic walk by cloning the ends of each YAC from lambda libraries. The left-hand ends of these YACs mapped to positions 115.2 and 193.5 in Figure 1. The right-hand ends of the 200- and 235-kilobase mouse inserts in YACs 3 and 4 should thus map approximately 70 and 185 kilobases past the edge of the earlier phage walk, respectively (see Figure 3). Probes isolated from the telomeric ends of both YACs were still deleted in DNA from the se^{20Zb} deletion, confirming the presence of a large deletion in this mutation. The chromosome walk was further expanded by an additional 27.4 kilobases past the edge of YAC4, and these probes were also deleted in se^{20Zb} DNA (Figure 4B).

Characteristics of Mice Carrying the 20Zb Deletion

Mice homozygous for the se^{20Zb} mutation have very short external ears and a spectrum of internal skeletal defects that resemble those reported in mice carrying the classical *short ear* allele (see, for example, the altered size and shape of the xiphoid process in Figure 5). Six of 11 se^{20Zb} homozygotes had 12 instead of 13 pairs of ribs,

FIGURE 5. Skeletal defects in mice carrying the se^{20Zb} mutation. Skeletons were prepared from 10-week-old male mice heterozygous or homozygous for the se^{20Zb} mutation. The animals shown are littermates, with the long-eared control animal on the left. The short-eared animal on the right has 12 instead of 13 pairs of ribs and a variety of subtle defects in the sizes and shapes of various bones and cartilage elements. Note the altered morphologies of the xiphoid process at the base of the sternum and the reduced size of the lateral processes on the lumbar vertebrae.

whereas 10 of 11 long-eared control littermates (+/se^{20Ab}) had 13 pairs of ribs. (The remaining heterozygous animal had 13 ribs on one side and 12 on the other.)

Test matings between heterozygous carriers of the se^{20Zb} mutation were set up to test whether homozygous absence of this large genomic region reduces the viability of mice. Five matings produced a total of 215 long-eared and 66 short-eared progeny. These numbers do not deviate significantly from the 3:1 ratio expected for segregation of a fully viable recessive mutation. Homozygous animals are fertile and have been used to maintain a continuous stock of mutant animals for over 30 years.

A Translocation Breakpoint in the T27H Mutation

Probes from the initial chromosome walk also identified a genetic alteration in a *short ear* mutation called T27H. This mutation was originally isolated as a new mutation at the *short ear* locus after irradiation of (C3H/HeH x 101/H)F_1 male mice with fission neutrons. Subsequent studies showed that the mutation is associated with a reciprocal chromosome translocation between the proximal portion of chromosome 1 and the distal portion of chromosome 9, which is lethal when homozygous, but produces viable short-eared mice in combination with either the classical *short ear* mutation or mutations like se^{20Zb} that carry large deletions on chromosome 9. These genetic studies, and the position of the cytological break, suggest that the chromosome 9 breakpoint is at the *short ear* locus. A 2.5-kilobase probe that begins at position 192.5 in Figure 1 detected a 7.2-kilobase *Bam*HI fragment in DNA samples from the various mouse strains used to generate and maintain the T27H mutation (C3H/R1, 101/R1, and C57BL/6 mice). This *Bam*HI fragment extends from positions 190 to 197.2 of the chromosome walk. In DNA samples from mice heterozygous for the T27H translocation, a larger *Bam*HI fragment was observed in addition to the wild-type fragment (Figure 2E). This larger fragment was not seen in the parental strains used to generate the T27H mutation, and therefore it must mark the position of at least one genetic alteration in the T27H mutation, possibly the position of the reciprocal translocation breakpoint itself.

The novel *Bam*HI fragment was cloned by screening T27H genomic DNA libraries with the 2.5-kilobase probe from position 192.5. Restriction mapping showed that the

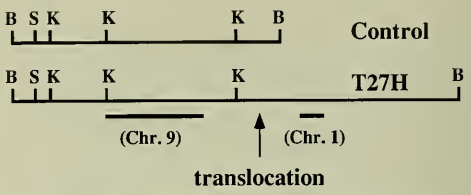


FIGURE 6. Restriction mapping of the T27H translocation breakpoint. The top line shows the 7.2-kilobase *Bam*HI fragment from wild-type DNA. This fragment extends from positions 190 to 197.2 in Figures 1 and 2. The bottom line shows the novel 12-kilobases *Bam*HI fragment cloned from T27H DNA. This fragment was isolated using the 2.5-kilobase probe from chromosome 9 and shows a number of restriction sites in common with the wild-type fragment (B, *Bam*HI; S, *Sal*I; K, *Kpn*I). The 0.5-kilobase *Hin*dII probe shown at the bottom right of the T27H clone has been mapped to chromosome 1 (see text). The translocation breakpoint in the T27H clone must occur somewhere between the *Kpn*I and *Bam*HI sites in the wild-type clone (arrow).

*Bam*HI fragment in T27H diverges from the wild-type *Bam*HI fragment at its distal end (Figure 6). A 500-base pair unique-sequence probe was isolated from the novel portion of the T27H fragment and mapped on an interspecific mouse cross that had previously been typed for over 1300 genetic markers distributed over all 19 autosomes and the X chromosome. The new probe was closely linked to the mouse *ST2* gene, which had previously been mapped to the proximal end of mouse chromosome 1. Four recombinants were seen between the T27H probe and the *ST2* gene in 70 animals, indicating that the two genes are 5.7 ± 2.8 cM apart. These data demonstrate that the altered *Bam*HI fragment contains the site of the (1:9) translocation breakpoint in the T27H mutation. Since the T27H mutation disrupts the *short ear* gene, we began a detailed search for possible *short ear* gene sequences in the region surrounding the translocation breakpoint, using an exon-trapping method.

Sequence of a Genomic Fragment Near the Translocation Breakpoint

In the course of exon-trapping experiments with a phage clone that extends from positions 176 to 195 in Figure 1, we cloned and sequenced a spliced product that we subsequently found to be the result of an aberrant splicing

event using a mouse splice acceptor sequence and a cryptic splice donor sequence located within the intron sequences of the splicing vector (data not shown). Similar splicing products have been reported by others using this technique. Although the sense strand of the mouse sequence did not detect any significant homologies in the DNA sequence databases, the opposite strand showed a perfect match to the first 34 base pairs at the 5′ end of a human cDNA clone for bone morphogenetic protein-5 (*BMP5*). Sequences that resemble a consensus splice acceptor sequence are also found on the antisense strand of the human sequence following position 34 and could explain the isolation of the original mouse clone. To confirm the homology with *BMP5*, we cloned and sequenced a larger genomic *Bgl*II fragment containing the original sequence. The extended mouse sequence is 90% identical to the 5′ end of a cDNA clone from human *BMP5* (Figure 7). The homology begins at the first base of the human sequence and continues until the end of the known mouse sequence. The evolutionary conservation in this region is particularly striking, since it is found in the 5′ untranslated region of the human sequence. Although different members of the *BMP* family are highly related to one another in many portions of their coding sequences, the 5′ untranslated region of the *BMP5* gene does not resemble that of the other *BMPs*. These data suggest that the 5′ end of the mouse *Bmp-5* gene is located within the 246-base pair genomic *Bgl*II fragment (positions 181.75 to 182 and gray boxes in Figures 1 and 3).

Mutations in the *Bmp-5* Gene in Different *Short Ear* Mutations

To determine the direction of transcription of the mouse *Bmp-5* gene, we isolated a probe that corresponds to the complete open reading frame of the human *BMP5* sequence. This probe hybridized to three additional regions of the various phage sets: one adjacent to the *Bgl*II fragment already sequenced (positions 182 to 183.5), one approximately 50 kilobases telomeric to this region (positions 233.6 to 234.2 in Figures 1 and 3), and one an additional 80 kilobases away in phage set B (positions 310 to 315 in Figure 3). In addition, a probe containing the C-terminal 187 base pairs of the human coding region hybridized to

FIGURE 7. Sequence comparison between a mouse genomic fragment and human *BMP5* cDNA. The sequence of a 246-base pair mouse genomic *Bgl*II fragment is shown on the upper line, with numbering beginning and ending at the *Bgl*II sites. The sequence of human *BMP5* is shown on the bottom line, with numbering beginning at the 5′ limit of the cDNA clone. Sequence identity is indicated by vertical lines. Three single nucleotide gaps have been introduced to maximize the alignment (shown by dashes). The coding region of the human *BMP5* gene does not begin until position 699 in the human sequence. The mouse fragment is located between 181.75 and 182 kilobases in Figures 1 and 3.

YAC4 DNA but not YAC3 DNA, suggesting that the highly conserved C terminus of the *Bmp-5* gene is located between phage sets B and C (Figure 3). These data demonstrate that the mouse *Bmp-5* gene is at least 135 kilobases long and is transcribed from left to right on Figures 1 and 3, which corresponds to centromeric to telomeric on chromosome 9. The coding regions of the *Bmp-5* gene map to both sides of the translocation breakpoint in the T27H mutation. The T27H mutation thus maps inside the *Bmp-5* gene and must interrupt the open reading frame of the gene.

The above studies suggest that most or all of the *Bmp-5* coding sequences should be located within the large deletion we have found in the se^{20Zb} mutation. The 1.3-kilobase *BMP5* coding region probe hybridized to seven *Eco*RI fragments in genomic DNA samples from wild-type mice (Figure 8). Six of these fragments were missing in DNA samples from mice homozygous for the se^{20Zb} mutation and from viable short-eared mice carrying both the *Da1* and *Db1* mutations. The single remaining *Eco*RI fragment in the mutant DNA samples hybridized less intensely than the other fragments seen in wild-type DNA samples and probably originated from cross-hybridization with other members of the *BMP* family.

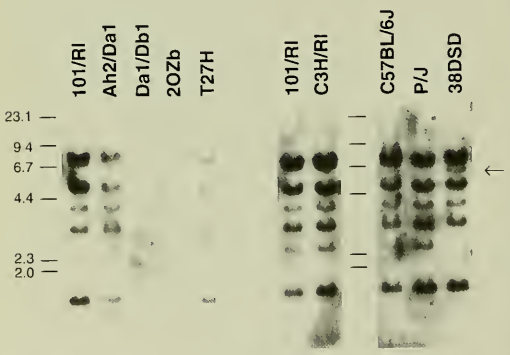


FIGURE 8. Deletions and rearrangements of *BMP5* coding sequences in *short ear* mutations. Ten micrograms of DNA from the indicated mouse strains was digested with *Eco*RI and analyzed by Southern blot hybridization with a 1.3-kilobase probe containing the complete open reading frame of the human *BMP5* gene. The 20Zb and 38DSD DNAs were isolated from short-eared mice homozygous for the se^{20Zb} and se^{38DSD} mutations. The T27H DNA was isolated from short-eared mice heterozygous for the T27H mutation and the se^{20Zb} deletion. The P/J DNA was isolated from a mouse strain homozygous for the classical *short ear* mutation. The other samples were as described in Figure 2. The arrow points to a unique *Eco*RI fragment seen in the se^{38DSD} mutation.

All of the restriction fragments recognized by the *BMP5* probe are present in DNA samples from mice carrying the T27H mutation in the presence of the se^{20Zb} mutation. These results demonstrate that the *Bmp-5* coding sequences are still present on the T27H-derived chromosomes, as expected for a reciprocal translocation. No rearrangements or deletions of *BMP5* sequences were observed in mice carrying the classical *short ear* mutation (a spontaneous mutation that has subsequently been incorporated into several different strains). Rearrangements were seen, however, in a viable radiation-induced *short ear* mutation called se^{38DSD}. In DNA samples from mice homozygous for the se^{38DSD} mutation, a 2.5-kilobase *Eco*RI restriction fragment recognized by the *BMP5* probe was missing and a novel fragment of approximately 6.7 kilobases appeared (Figure 8). Altered restriction fragments were also detected with three additional enzymes (data not shown). The se^{28DSD} mutation was originally generated by treatment of F_1 hybrids between C3H/Rl and 101/Rl mice with 100 + 500 rads of fractionated acute X irradiation (24 hours between fractions). None of the novel fragments is seen in control DNA samples from the original C3H/Rl or 101/Rl strains. In addition, no deletions or rearrangements in this mutation were detectable using 20 different chromosome walking probes from phage sets A, B, C, and D in Figure 2, including probes that map upstream and downstream of the *Bmp-5* gene. The altered fragments are thus likely to have arisen from an induced mutation in the *Bmp-5* gene.

Thus, the current data show that the mouse *Bmp-5* gene is located at or near the *short ear* locus. This gene maps inside a large interval that must contain the *short ear* gene as defined by the large deletions present in three different multigenic mutations (*Da1, Db1, Ah2*) and one viable allele at the *short ear* locus (se^{20Zb}). In addition, a portion of the *Bmp-5* gene is rearranged in the se^{38DSD} mutation, and the open reading frame of the gene is interrupted by a reciprocal chromosome translocation in the T27H mutation. These genetic alterations and the known biochemical properties of bone morphogenetic proteins provide strong evidence that the *short ear* locus encodes *Bmp-5*.

Formal proof of the identity between the two genes will require the functional complementation of the mutant phenotypes seen in *short ear* mice with cloned *Bmp-5* sequences.

BMPs and the TGF-β Superfamily

The *BMP5* gene product is a member of a large superfamily of proteins related to transforming growth factor β (TGF-β). In recent years, it has become clear that the various members of this family are strikingly conserved during evolution and play diverse and important roles in many different organisms. The prototype of the family, *TGF-β1*, was originally identified as a factor that promoted the growth in soft agar of a rat kidney fibroblast cell line. Subsequent studies have shown that this factor inhibits rather than stimulates the growth of many other cell lines and has a wide number of effects on differentiation and matrix production. In mammals, there are at least a dozen genes that are clearly related to *TGF-β1*. *TGF-β2* and *-β3* are structurally very similar to *TGF-β1* and have similar activities. Activins and inhibins control the production of various gonadotropic hormones by pituitary and placental cells and can also modulate the differentiation of erythroleukemia cells. Mullerian inhibiting substance causes the regression of female sex organ precursors in males. Bone morphogenetic proteins induce the aggregation, proliferation, and differentiation of mesenchyme cells when implanted under the skin or into the muscles of animals. This process leads to a series of histological transformations that closely resemble those seen in normal embryonic bone formation, and ultimately results in the formation of ectopic cartilage and bones at the site of implantation.

The various members of the TGF-β family appear to be synthesized as secreted precursor proteins with an N-terminal signal sequence and protease domain. Proteolytic processing of these precursors releases a highly conserved C-terminal domain consisting of approximately 110 amino acids. This domain contains seven highly conserved cysteine residues and forms homo- or heterodimers with other subunits. Regulation of activity can occur at many levels, including proteolytic processing, release of active factor from an inactive complex, and association of different subunits to produce signaling molecules with different activities.

The bone morphogenetic proteins represent a distinct subset of the larger TGF-β superfamily. In mammals, a total of seven different *BMPs* have been identified based on their ability to induce the formation of ectopic bone and cartilage when implanted under the skin of rats or by their homology to such factors. Six of the seven mammalian *BMPs* are members of the TGF-β superfamily, and the remaining *BMP* includes a protease domain that may activate the other factors. Mammalian *BMP2* and *BMP4* are more closely related to the *Drosophila dpp* gene than they are to the other mammalian *BMPs*, suggesting that these genes represent the vertebrate homologs of the *dpp* gene. Similarly, the *BMP5*, *-6*, and *-7* genes are more closely related to a recently discovered *Drosophila* gene called *60A* than to either *dpp* or the other *BMPs*, suggesting that they represent the vertebrate homologs of the *60A* gene. These data suggest that the bone morphogenetic proteins can be subdivided into at least two classes and that these classes have existed at least since the time of the last common ancestor of arthropods and chordates (over 500 million years ago).

Function of *BMPs* in Development

The striking evolutionary conservation of the *BMP* genes suggests that they play critical roles in the normal development or function of animals. The existing biochemical experiments with BMPs in mammals show that, individually or in combination, these factors can induce the complex series of chemotactic, aggregation, and differentiation responses required to produce cartilage and bone in vivo. These observations have led to speculation that these genes may normally control the formation of skeletal tissue during embryogenesis. Results of in situ hybridization studies are consistent with a role for *BMPs* in normal skeletal development but also suggest that these genes are expressed in many other tissues. The mouse *Bmp-2* gene, for example, is expressed in the apical ectodermal ridge of developing limb buds, and in condensing precartilaginous mesenchyme of vertebrae, but is also expressed in hair and whisker follicles, myogenic layers of the heart, and the otic vesicle. The mouse *Bmp-6* gene (also known as *Vgr-1*) is expressed in hypertrophic cartilage of developing limbs but is also expressed in many different stratified squamous epithelia, the developing nervous system, oocytes, and many adult tissues.

No mutations have been available to test the functions of any of the *BMPs* in vivo. The finding that *Bmp-5* coding sequences are extensively deleted in *short ear* mutations thus provides the first genetic test of the function of these genes in higher animals. It is striking that the large deletion we have demonstrated in the se^{20Zb} mutation indicates that this mutation represents a null allele at the *Bmp-5* locus.

The phenotypes seen in various mutations at the *short ear* locus suggest that bone morphogenetic proteins have been aptly named. The predominant phenotype in short-eared mice is a specific syndrome of morphological alterations in numerous skeletal elements. These alterations affect the presence or absence of many small bones, the shape of other bones such as the sternum, the presence or absence of specific processes on bones such as the sixth cervical vertebrae, and the number of other repeating skeletal elements such as ribs. Where they have been studied in detail, these defects are known to result from defects in the earliest condensation stage of skeletal development, prior to the overt differentiation of bone and cartilage. Specific condensations in short-eared animals have altered shapes and sizes, suggesting that the gene causes local defects in the production or interpretation of the signals that cause cells to aggregate into the outlines of particular skeletal elements. BMP preparations are known to induce condensations of mesenchyme cells when implanted in ectopic sites under the skin of adult animals. The defects of short-eared animals suggest that BMPs are the natural inducers of similar condensations during embryonic development.

Many of the morphological features of normal bone development can also be observed in adult animals during the repair of bone fractures. Purified preparations of BMPs are known to stimulate the repair of bone fractures in adult animals and may be used clinically for this purpose. Previous studies have shown that *short ear* mice have a reduced capacity to repair rib fractures as adults, suggesting that the molecular mechanisms used to induce bone formation in embryonic development and in fracture repair are similar. The current data suggest that *Bmp-5* plays an important role in both processes.

A number of soft-tissue abnormalities have also been described in *short ear* mice, including an increased frequency of misplaced ovaries, hydrotic kidneys, lung cysts, liver granulomas, and neuromuscular tail kinks. In the past, it was not clear whether such defects were secondary consequences of skeletal defects in short-eared animals, or whether they reflect a requirement for the *short ear* gene. However, the closely related *Bmp-6* and *-7* genes have been shown to be expressed in many extraskeletal sites including kidneys, lungs, and liver. It will be interesting to determine whether the soft-tissue defects in short-eared animals reflect additional roles for bone morphogenetic proteins in normal development.

Genetic Control of Skeletal Morphology in Vertebrates

It is important to note that the various skeletal alterations seen in short-eared animals occur against a background of an otherwise normal skeleton. These data suggest that the *Bmp-5* gene is required only for specific aspects of skeletal formation and patterning. Other *BMPs* may play important roles in the formation of other skeletal structures, including the many bones that are not significantly altered in *short ear* mice. Studies with cloned BMPs have shown that individual BMPs can induce bone formation when tested in the ectopic bone formation assay. The family of related *BMP* genes may thus provide an overlapping set of signals that can induce bone formation in different body regions.

It is striking that many of the morphological traits affected by the *short ear* gene are the same sorts of traits that vary among different organisms (size and shape of the external ear, the number of ribs along the vertebral column, presence or absence of particular processes on different bones). The specific skeletal defects in mice carrying the *short ear* mutation suggest that complicated morphological traits can be controlled by the expression of individual members of the BMP family. Many evolutionary specializations in skeletal form and pattern may arise through genetic variation in the sites of expression, dimerization, and processing of other BMPs.

Additional studies of *short-ear* mutations may help to identify how particular aspects of morphology are specified by *Bmp-5*. Over two dozen spontaneous, radiation-induced, and chemical-induced mutations have previously been isolated at the *short ear* locus. Many of these mutations have slightly different effects on the morphological features seen in short-eared animals. In addition, the classical *short*

ear mutation has already been introduced into several different genetic backgrounds by years of breeding. Interestingly, the phenotypic effects of this single mutation are strongly dependent on various modifier genes present in different mouse strains. In *Drosophila*, similar modifier genes have helped to identify factors that act upstream and downstream of the *BMP* homolog, *dpp*. The mouse modifier genes can now be reexamined for their possible effects on the production and distribution of BMPs or on the cellular response to such signals. Genetic variation in the BMP signaling pathway may be responsible for many different aspects of morphological variation and skeletal diseases in higher organisms.

Joint Projects Between the Molecular Genetics of Oncogenesis Section and the Molecular Genetics of Development Section

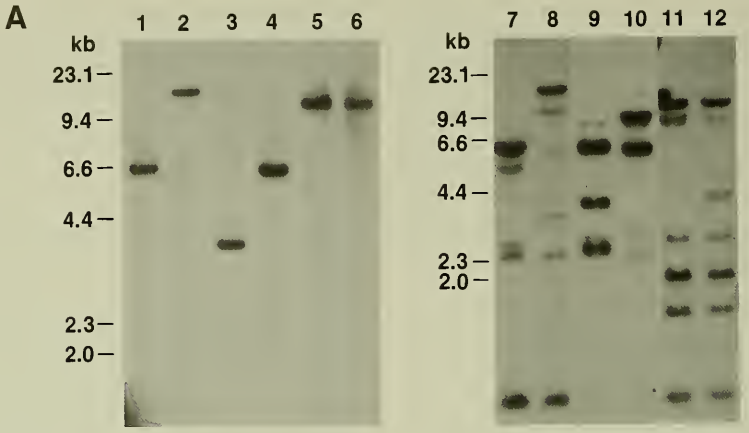
Over the past few years, we have used interspecific mouse backcrosses (IB) to develop a molecular genetic linkage map of the mouse genome. More than 1300 loci have been placed on the map; the loci are fairly well distributed over all the mouse autosomes and the X chromosome. The average distance between loci is ~1.2 cM. With this level of resolution, the probability of being able to position a new gene on the map is 100%. What follows is a summary of the major applications of the map that were published during the preceding year. We gratefully acknowledge the many collaborators in these studies.

Maternal Imprinting of the Mouse *Snrpn* Gene and Conserved Linkage Homology with the Prader-Willi Syndrome Region of Humans

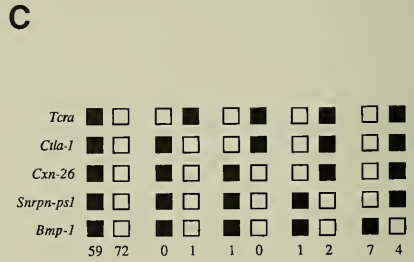
The Sm autoantigens are core proteins of small nuclear ribonucleoproteins (snRNPs) that are involved in mRNA splicing. These antigens are recognized by anti-Sm antibodies that are commonly found in the sera of systemic lupus erythematosus patients. Using anti-Sm sera, a tissue-specific snRNP from rat called SmN that shared an epitope common to other Sm antigens was cloned. SmN mRNA is most abundant in brain, less so in heart, and barely detectable in all other tissues examined. Consequently, it has been suggested that SmN may be involved in regulating tissue-specific splicing events. As a first step in determining whether *Snrpn* is associated with a known mouse mutation or a human genetic disease, we mapped its chromosomal location on high-resolution mouse linkage and human-mouse comparative maps.

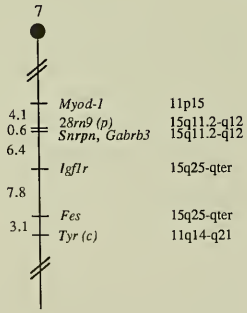
In initial mapping studies, we used an *Snrpn* probe obtained by PCR amplification of mouse genomic DNA with oligonucleotides predicted to span intron 5. We obtained a 750-base pair fragment that detected only one band on Southern blots with a majority of enzymes tested (Figure 9A, lanes 1 to 6). We then determined the chromosomal location of *Snrpn* by IB analysis using progeny derived from mating of [(C57BL/6J x *Mus spretus*)F_1 x C57BL/6J] mice. This IB mapping panel has been typed for over 1300 loci that are well distributed among the autosomes as well as the X chromosome. C57BL/6J and *M. spretus* DNAs were digested with several enzymes and analyzed by Southern blot hybridization for informative restriction fragment length polymorphisms (RFLPs) using the *Snrpn* intron 5 probe. A 13.6-kilobase *M. spretus*-specific *Taq*I RFLP was used to follow the segregation of the *Snrpn* locus in backcross mice. The mapping results indicated that *Snrpn* is located in the middle region of mouse chromosome 7 linked to *Myod-1, 28rn9 (p), Gabrb3, Igf1r, Fes,* and *Tyr(c)* (Figure 9B).

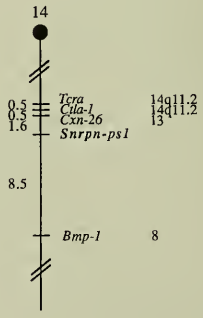
We also used a 1.3-kilobase *Snrpn* murine cDNA probe in the mapping studies. While the intron 5-specific probe detected only one band on Southern blots that mapped to chromosome 7, the 1.3-kilobase *Snrpn* cDNA probe identified two independently segregating loci in the backcross panel. One of these loci mapped to chromosome 7, identical in position to the locus identified with the intron 5-specific probe, while the second locus mapped to chromosome 14 linked to *Tcra, Ctla-1, Cxn-26,* and *Bmp-1* (Figure 9C). Additional experiments suggested that the chromosome 14 locus represents a nonexpressed pseudogene. First, PCR amplification of mouse and mouse x Chinese hamster hybrid cell genomic DNAs with primers from sequences predicted to encode exons 2 and 3 of the *Snrpn* gene indicated that the chromosome 7 locus contains intron 2 while the chromosome 14 locus lacks this intron (Figure 10). Second, PCR analysis of the chromosome 14 locus using primers complementary to sequences


A
1 2 3 4 5 6
kb
23.1
9.4
6.6
4.4
2.3
2.0
7 8 9 10 11 12
kb
23.1
9.4
6.6
4.4
2.3
2.0


B
Myod-1
28rn9
Snrpn, Gabrb3
Igf1r
Fes
Tyr
61 55 1 5 1 0 4 5 3 6 1 3


C
Tcra
Ctla-1
Cxn-26
Snrpn-ps1
Bmp-1
59 72 0 1 1 0 1 2 7 4


7
Myod-1 11p15
4.1
28rn9 (p) 15q11.2-q12
0.6
Snrpn, Gabrb3 15q11.2-q12
6.4
Igf1r 15q25-qter
7.8
Fes 15q25-qter
3.1
Tyr (c) 11q14-q21


14
0.5
Tcra 14q11.2
0.5
Ctla-1 14q11.2
1.6
Cxn-26 13
Snrpn-ps1
8.5
Bmp-1 8

predicted to encode exons 1 and 6 of the *Snrpn* gene showed the complete absence of introns in the chromosome 14 locus, several nucleotide and amino acid differences from the SmN cDNA, and interruption of the SmN open reading frame by a stop codon at amino acid 32 (data not shown). Third, sequencing of a number of reverse transcriptase (RT)-PCR products of C57BL/6J brain RNA showed that only the chromosome 7 locus is expressed (data not shown).

The mapping of *Snrpn* to a region of chromosome 7 containing *Gabrb3* and *p* raised the interesting possibility that *Snrpn* might be located in the Prader-Willi Syndrome (PWS) and Angelman Syndrome (AS) region of human chromosome 15q11-q13. Probe 28rn9 from the *p* locus has been shown to be very closely linked in the mouse to probe DN10 using RI lines, and recent studies have shown that DN10 is in fact a human homolog of the mouse *p* gene. Both DN10 and *FABRB3* have been demonstrated to map within some PWS/AS deletions. Indeed, in a separate study, it was shown that the human *SNRPN* gene does map within the minimally deleted region for PWS, but not for AS.

The discovery that PWS could arise from the inheritance of two copies of a maternal chromosome and none from the father provided an important insight into the genetics of PWS. The observation of maternal disomy for chromosome 15q11-q13 supported the hypothesis that PWS may be caused by the lack of one or more paternally expressed genes within 15q11-q13 with the corresponding maternally inherited copies inactivated through imprinting. To determine whether *Snrpn* is maternally imprinted in the mouse, we used a previously developed imprinting assay. Specifically, an RNase protection assay that can detect subtle base substitutions, deletions, or insertions was used to discriminate between SmN transcripts arising from *Mus musculus* (C57BL/6J) and *M. spretus* alleles. An antisense RNA probe that included 33 nucleotides of N-terminal-coding and 446 nucleotides of 5′ noncoding sequence of an SmN cDNA (Figure 11A) was used in these studies. The probe was hybridized to total brain RNA from C57BL/6J and *M. spretus* or their

FIGURE 9. (opposite page) Mapping the *Snrpn* gene by interspecific backcross analysis. (A) Southern blot analysis using the *Snrpn* intron 5 probe (lanes 1 to 6) or a 1.3-kilobase *Snrpn* cDNA probe (lanes 7 to12). Even-numbered lanes contain DNA derived from C57BL/6J animals and odd-numbered lanes contain *M. spretus* DNA. DNA samples in lanes 1, 2, 7, and 8 were digested with *Taq*I; lanes 3, 4, 9, and 10 with *Bgl*II; and lanes 5, 6, 11, and 12 with *Pst*I. (B) *Snrpn* maps in the proximal region of mouse chromosome 7. The segregation patterns of *Snrpn* and flanking genes in 134 backcross animals that were typed for all loci are shown at the top of the figure. Each column represents the chromosome identified in the backcross progeny that was inherited from the (C57BL/6J x *M. spretus*) F_1 parent. Shaded boxes represent the presence of a C57BL/6J allele and white boxes represent the presence of a *M. spretus* allele. The number of offspring inheriting each type of chromosome is listed at the bottom of each column. Although 134 mice were analyzed for every marker and are shown in the segregation analysis, up to 193 mice were typed for some pairs of markers. Each locus was analyzed in pairwise combinations for recombination frequencies using the additional data. The ratios of the total number of mice exhibiting recombinant chromosomes to the total number of mice analyzed for each pair of loci and the most likely gene order are: centromere - *Myod-1* - 6/147 - *28rn9 (p)* - 1/164 - *Gabrb3* - 0/183 - *Snrpn* - 12/188 - *Igf1r* - 15/193 - *Fes* - 6/193 - Tyr (c). The recombination frequencies [expressed as genetic distances in centimorgans (cM) ± the standard error] are *Myod-1* - 4.1 ± 1.6 - *28rn9 (p)* - 0.6 ± 0.6 - [*Gabrb3*, *Snrpn*] - 6.4 ± 1.8 - *Igf1r* - 7.8 ± 1.9 - *Fes* - 3.1 ± 1.3 - *Tyr* (c). No recombinants were detected between *Gabrb3* and *Snrpn* in 183 animals typed in common, suggesting that the two loci are within 1.6 cM of each other (upper 95% confidence limit). A partial chromosome 7 linkage map showing the location of *Snrpn* in relation to linked genes is shown underneath the segregation patterns. Recombination distances between loci in centimorgans are shown to the left of the chromosome and the positions of loci in human chromosomes, where known, are shown to the right. (C) *Snrpn-ps1* maps to chromosome 14. *Snrpn-ps1* was placed on mouse chromosome 14 by interspecific backcross analysis using a 1.3-kilobase *Snrpn* murine cDNA probe. The segregation patterns of *Snrpn-ps1* and flanking genes in 147 backcross animals that were typed for all loci are shown at the top of the figure. The 8.7-kilobase *M. spretus*-specific *Bgl*II RFLP was used to follow the segregation of the second locus, designated *Snrpn-ps1*, in backcross mice. Although 147 mice were analyzed for every marker and are shown in the segregation analysis, up to 197 mice were typed for some pairs of markers. The ratios of the total number of mice exhibiting recombinant chromosomes to the total number of mice analyzed for each pair of loci and the most likely gene order are: centromere - *Tcra* - 1/197 - *Ctla-1* - 1/188 - *Cxn-26* - 3/184 - *Snrpnps1* - 13/153 - *Bmp-1*. The recombination frequencies are *Tcra* - 0.5 ± 0.5 - *Ctla-1* - 0.5 ± 0.5 - *Cxn-26* - 1.6 ± 0.9 - *Snrpn-ps1* - 8.5 ± 2.3 - *Bmp-1*. A partial chromosome 14 linkage map showing the location of *Snrpn-ps1* in relation to linked genes is shown underneath the segregation patterns. References for the human map positions of loci mapped in this study can be obtained from GDB (Genome Data Base), a computerized database of human linkage information maintained by the William H. Welch Medical Library of the Johns Hopkins University.

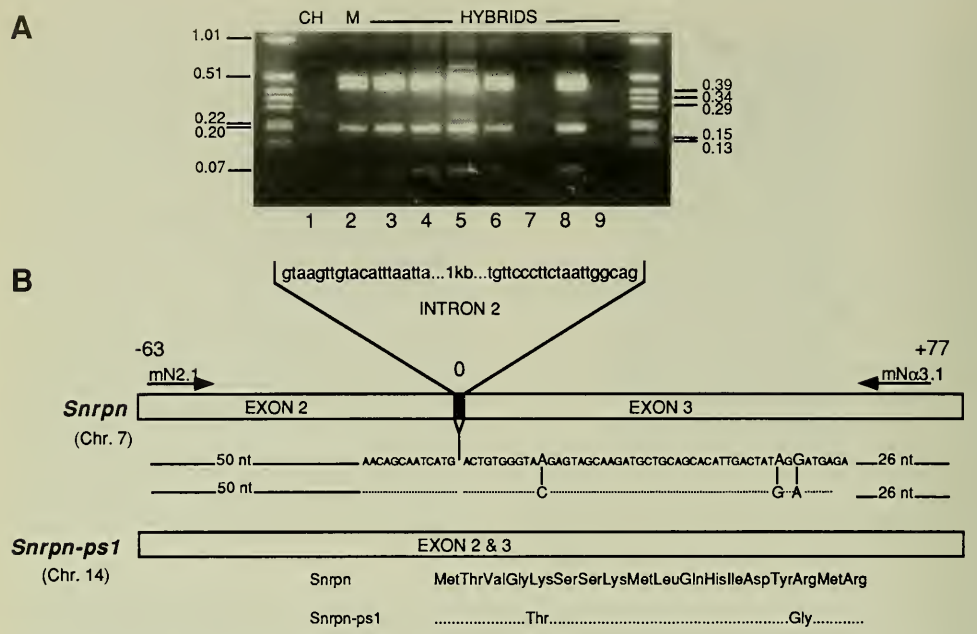


FIGURE 10. Genomic sequence of the chromosome 7 (*Snrpn*) and 14 (*Snrpn-ps1*)–specific genes surrounding intron 2. (A) An agarose-ethidium bromide gel showing *RsaI* digestion products of a 1.2-kilobase PCR product amplified from mouse x Chinese hamster hybrid cell DNAs using primers mN2.1 and mNα3.1. *RsaI* digests distinguish hamster (lane 1) and mouse (lane 2) DNA. Analysis of hybrid cell DNAs that have defined subsets of mouse chromosomes revealed concordance of the mouse-specific fragments with chromosome 7 and discordance with chromosome 14. Hybrids in lanes 3 to 6 and 8 are positive and those in lanes 7 and 9 are negative for the mouse-specific fragments and for the presence of chromosome 7. Flanking lanes contain 1-kilobase ladder DNA as marker. (B) The chromosome 7-specific 1.2-kilobase PCR product (upper) contains portions of exons 2 and 3 separated by a 1.05-kilobase intron (intron 2, black box). The 140-base pair chromosome 14-specific PCR product contains sequences homologous to portions of exons 2 and 3 of the chromosome 7-specific *Snrpn* gene fused together without intron 2. Nucleotide differences between *Snrpn* and *Snrpn-ps1* are indicated by bars. These nucleotide substitutions predict two differences in the first 27 amino acids of these two potential gene products, as indicated in the two lower lines.

progeny from a number of F_1 or N_2 crosses. After hybridization, the RNase-resistant products were analyzed by polyacrylamide gel electrophoresis. C57BL/6J RNA protected the expected 480-nucleotide fragment (Figure 11B, lanes 2 and 12). *M. spretus* RNA protected a smaller fragment, approximately 60 nucleotides less than that protected by C57BL/6J RNA (Figure 11B, lanes 3 and 13). When brain RNAs from the F_1 progeny of C57BL/6J females mated with *M. spretus* males were tested with the RNase protection assay, only the *Snrpn* gene product characteristic of the male *M. spretus* parent was detected (Figure 11B, lanes 4 to 7). Conversely, in the reciprocal F_1 progeny, we found only the paternal C57BL/6J allele of the *Snrpn* gene expressed in RNA isolated from heads of neonates (Figure 11B, lanes 8 and 9) or the brain of a 6-week-old mouse (data not shown).

If the *Snrpn* gene is indeed maternally imprinted as these studies suggest, the pattern of allele-specific expression will be reset in the germ line at each generation. To test this prediction, female (C57BL/6J x *M. spretus*)F_1

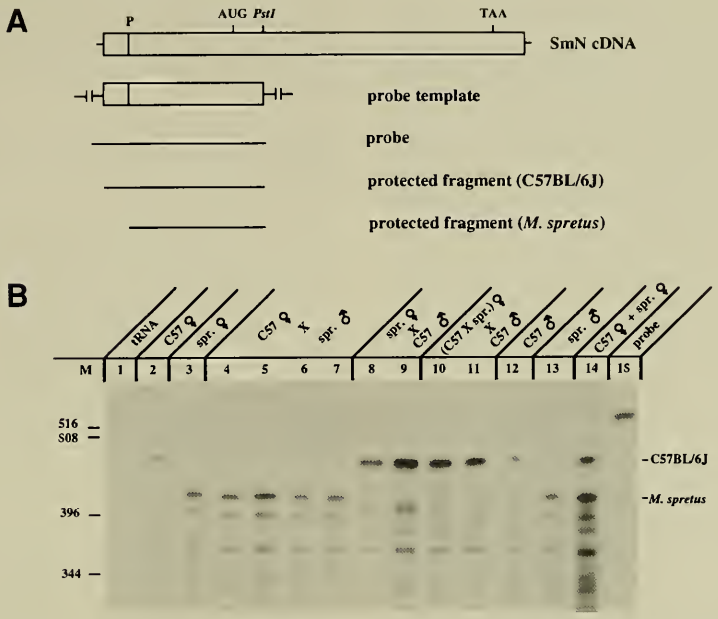


FIGURE 11. RNase protection of brain RNA from C57BL/6J, *M. spretus*, and hybrid progeny. (A) The mouse SmN cDNA is depicted in the first line and the second line shows the 5′ region that was subcloned to use as a template for an RNase protection probe. The subclone bounded at the 3′ end by a *Pst*I site, contained 478 nucleotides of cDNA sequence (indicated by the box), which included the first 33 nucleotides of N-terminal coding sequence. P indicates the predicted site of a polymorphism between the C57BL/6J and the *M. spretus* alleles. The broken lines represent the vector sequences. A 538-base pair antisense RNA (third line) was generated and used as a probe in the RNase protection assay. The fourth and fifth lines represent the fragments protected by digestion by RNase after hybridization with C57BL/6J or *M. spretus*, respectively. (B) Autoradiograph of RNase protection assays of 10 μg of adult brain RNA (lanes 2 to 7 and 10 to 14) or 40 μg of neonatal head RNA (lanes 8 and 9). RNAs from parental strains show a major protected band at approximately 480 nucleotides for C57BL/6J (lanes 2 and 12) and 420 nucleotides for *M. spretus* males (lanes 4 to 7); only the 420-nucleotide product characteristic of *M. spretus* was detected. Using RNA from the F_1 progeny of *M. spretus* females mated to C57BL/6J males, only the 480-nucleotide product was detected (lanes 8 and 9). Lanes 10 and 11 show the results of RNA from N_2 progeny of a backcross between female (C57BL/6J x *M. spretus*)F_1 and a C57BL/6J male. Lane 14 shows an assay control using equal amounts of C57BL/6J and *M. spretus*. (As both protected fragments are equally intense, the probe is able to detect both types of RNA.) A negative tRNA control is shown in lane 1 and undigested probe is shown in lane 15. The size markers (M) were generated by T4 polynucleotide kinase treatment of *Hin*fI-digested pBR322 DNA.

mice expressing only the *M. spretus* SmN allele from the F_1 female parent were identified by *Taq*I RFLP analysis of backcross DNAs (data not shown), and their brain RNA was assayed by RNase protection. In a total of 6 backcross progeny tested, all displayed exclusive expression of the paternal C57BL/6J allele (Figure 11B, lanes 10 and 11), demonstrating that the formerly active *M. spretus* allele was now inactive. We conclude, therefore, that *Snrpn* is subject to maternal imprinting in the mouse such that only the paternally inherited copy of *Snrpn* is expressed. This imprint direction is consistent with a gene involved in PWS. Additional studies will be required to determine whether *Snrpn* is maternally imprinted in humans.

PWS and AS involve different phenotypes arising from paternal versus maternal inheritance of genetic defects on human chromosome 15q11-q13. Recent studies have suggested that although the genetic loci for PWS and AS are closely linked, they are nonetheless distinct. A model to explain the differential imprinting of the two closely linked genes, *Igf2* and *H19*, has been proposed. The maternally imprinted *Igf2* gene and the paternally imprinted *H19* gene may compete for a common set of regulatory elements such that only one gene is active per chromosome. The successful competition for the common enhancer would be biased in opposite directions on the two chromosomes by differential epigenetic markings. It will be interesting to determine whether *Snrpn* is similarly coupled to a gene involved in AS.

What role could *Snrpn* play in the phenotypes associated with PWS? Among the common clinical features of PWS are grossly diminished fetal activity, severe infant hypotonia, hypogonadism and hypogenitalism, poor feeding ability in infancy with later insatiable appetite leading to obesity, retarded bone development and short stature, small hands and feet, delayed mental and psychomotor development, and characteristic facies. These features can be quite variable among individual patients, suggesting that multiple genes are involved in PWS. Although SmN is associated with the RNA splicing machinery of neural cells, the precise biochemical function of SmN protein in mRNA processing remains unclear. SmN has been postulated to function as a trans-acting factor promoting calcitonin gene-related peptide (CGRP) mRNA production. CGRP is a brain-specific alternately spliced product of the calcitonin gene that functions as a hypothalamic peptide and is reported to be active in inhibiting feeding behavior in rats. CGRP is also believed to regulate cardiovascular and gastrointestinal homeostasis. So far, experiments involving transfection of the *Snrpn* gene in tissue culture have failed to result in CGRP production. It still remains possible, however, that the high levels of SmN protein found in the brain are required for efficient CGRP production. Alternatively, it is possible that SmN acts as a regulator of splicing efficiency or alternative splicing of a variety of genes and that the pleiotropic effects observed in PWS may simply be due to the lack of such a master regulator. Using homologous recombination in embryonic stem (ES) cells to generate a null allele for *Snrpn* in mice, we hope to answer these questions. This work was performed in collaboration with Stewart E. Leff, Martha Reed, Tayfun Ozcelik, and Uta Franke (Stanford University School of Medicine).

Molecular Cloning of the Murine IL-1β–Converting Enzyme cDNA

Interleukin-1β (IL-1β) is a potent modulator of immune and inflammatory responses. Murine IL-1β is initially synthesized as an inactive 33-kilodalton pro-molecule that is activated by proteolytic cleavage between Asp117 and Val118 to generate the 17-kilodalton mature IL-1β protein. This cleavage is catalyzed by a specific protease that has been designated the IL-1β–converting enzyme (or IL-1β convertase). A human IL-1β convertase cDNA was used to isolate a murine convertase cDNA from a WEHI-3 library. Southern blot analysis of interspecific backcross mice indicated that the murine IL-1β convertase is encoded by a single-copy gene located on murine chromosome 9.

Characterization of the Human Pleiotrophin Gene

The protein encoded by the pleiotrophin (*PTN*) gene belongs to a recently described family of heparin-binding cytokines whose expression is temporally and spatially regulated during development. Using a probe to the human *PTN* gene, the gene was localized to human chromosome 7, band q33, and mouse chromosome 6, respectively.

Cloning, Chromosomal Localization, and Expression Pattern of the POU Domain Gene *Oct-11*

POU domain genes encode a family of highly conserved trans-acting factors that influence the transcriptional activity of several cell-type-specific and ubiquitous genes. A novel mouse POU domain protein, Oct-11, that is closely related within the POU domain to the POU class II proteins, Oct-1 and Oct-2, was cloned and sequenced. Recombinant Oct-11 protein bound specifically to an octamer sequence in vitro. The *Oct-1* gene was expressed during mouse embryogenesis and in the adult thymus and testis. In addition, it was abundant in the myeloma cell line P3/NS-1/1-Ag4.1. The *Oct-11* structural gene was mapped to mouse chromosome 1 by IB analysis.

Expression of a Novel Form of *Tec* Kinase in Hematopoietic Cells and Mapping of the Gene to Chromosome 5 Near *Kit*

The *Tec* kinase was initially identified as a novel cytoplasmic protein tyrosine kinase that is preferentially expressed in the liver and is highly homologous to the *Drosophila Dsrc*28C Src-related tyrosine kinase. In the screening of interleukin-3 (IL-3)-dependent myeloid leukemia cells for protein tyrosine kinases, it was found that all cell lines examined expressed high levels of *Tec* transcripts. However, characterization of *Tec* cDNAs indicated that they differed significantly from the published sequence. Most strikingly, an insertion of 41 base pairs in the 5′ region affects the initiation codon and results in replacing the published 13-amino acid N-terminal sequences with 94 amino acids. By PCR analysis, only the form containing the insertion was detected in hematopoietic cells. In addition, an in-frame insertion of 66 base pairs that introduces an additional 22 amino acids into the SH3 domain was found. This insertion restores conserved SH3 sequences that are found in the *src* gene family and in the D*src*28C gene. By PCR analysis, approximately equal levels of *Tec* transcripts containing the intact SH3 domain and the 22-amino acid deletion were found in hematopoietic cells. Lastly, by IB analysis, we showed that the *Tec* gene is tightly linked to the c-*kit* gene on mouse chromosome 5.

Anchored Reference Loci for Comparative Genome Mapping in Mammals

Recent advances in mapping technologies have led to increased emphasis in developing representative genetic maps for several species, particularly domestic plants and animals. These maps are being compiled with two distinct goals: to provide a resource for genetic analysis, and to help dissect the evolution of genome organization by comparing linkage relationships of homologous genes. We proposed a list of 321 reference anchor loci suitable for comparative gene mapping in mammals and other vertebrate classes. We selected cloned mouse and human functional genes spaced an average of 5 to10 cM throughout their respective genomes. We also attempted to include loci that are evolutionarily conserved and well represented in comparative gene maps in other mammalian orders, particularly cattle and the domestic cat. We believe that the map may provide the basis for a unified approach to comparative analysis of mammalian species genomes.

Transgenic Mouse Model of X-Linked Cleft Palate

A transgenic mouse line, PyLMP.5, exhibited a sex-linked lethality not observed in any other lines expressing the transgene. In this unique line, the transgene integrated into the X chromosome, yielding a simple tandem duplication of the insert sequences with minimal, if any, additional rearrangement of the cellular sequences. The predominant phenotype was a cleft secondary palate and neonatal lethality in males. Survival of females was dependent on the mouse strain background. The disrupted cellular sequences were mapped to the proximal region of the mouse X chromosome. The disrupted locus may represent the mouse counterpart to a human locus mutated in an X-linked cleft secondary palate syndrome.

Chromosomal Localization of Mouse Bullous Pemphigoid Antigens, BPAG1 and BPAG2: Identification of a New Region of Homology Between Mouse and Human Chromosomes

Two bullous pemphigoid antigens, BPAG1 and BPAG2, have been recently cloned and mapped to human chromosomes 6p12-p11 and 10q24.3, respectively. We localized the corresponding mouse genes by IB analysis. *Bpag-1* mapped to the proximal region of mouse chromosome 1, identifying a new region of homology between human chromosome 6 and mouse chromosome 1. *Bpag-2* mapped to the distal end of mouse chromosome 19 in a region of homology to human chromosome 10q. These assignments confirmed and extended the relationships between the human and mouse chromosomes.

CD40 Ligand Gene Defects Responsible for X-Linked Hyper-IgM Syndrome

The ligand for CD40 (CD40L) is a membrane glycoprotein on activated T cells that induces B-cell proliferation and immunoglobulin secretion. Abnormalities in the

CD40L gene were associated with an X-linked immunodeficiency in humans [hyper-IgM (immunoglobulin M) syndrome]. This disease is characterized by elevated concentrations of serum IgM and decreased amounts of all other isotypes. CD40L cDNAs from three of four patients with this syndrome contained distinct point mutations. Recombinant expression of two of the mutant CD40L cDNAs resulted in proteins incapable of binding to CD40 and unable to induce proliferation or IgE secretion from normal B cells. Activated T cells from the four affected patients failed to express wild-type CD40L, although their B cells responded normally to wild-type CD40L. Thus, these CD40L defects lead to a T-cell abnormality that results in the failure of patient B cells to undergo immunoglobulin class switching.

Cloning, Analysis, and Chromosomal Localization of *Notch-1*, a Mouse Homolog of *Drosophila Notch*

The *Notch* gene of *Drosophila* encodes a large transmembrane protein involved in cell-cell interactions and cell fate decisions in the *Drosophila* embryo. cDNA clones encompassing the full-length coding sequence of *Notch-1*, a mouse homolog of *Drosophila Notch*, were isolated. The predicted amino acid sequence of the Notch-1 protein retains all of the conserved amino acid motifs of Notch and the other vertebrate Notch homologs. The cDNA sequence predicts a 2531-amino acid protein containing a signal peptide, 36 epidermal growth factor-like repeats, 3 *Notch-lin-12* repeats, a transmembrane domain, and 6 *cdc10/ankyrin* repeats. The *Notch-1* gene was localized to the proximal portion of mouse chromosome 2 by mapping with an IB panel.

Localization of the 70-Kilodalton Peroxisomal Membrane Protein to Human 1p21-p22 and Mouse 3

The 70-kilodalton peroxisomal membrane protein (PXMP1) is a member of the ATP-binding cassette transporter family. In humans, mutations in this gene may be responsible for a subset of patients with Zellweger Syndrome, a lethal inborn error of peroxisome assembly. The *PXMP1* gene was assigned to human chromosome 1p21-p22 by in situ hybridization, and its murine homolog (*Pxmp-1*) was assigned to chromosome 3 by IB analysis.

Molecular Characterization and Mapping of Mu Genes Encoding Three Members of the Stefin Family of Cysteine Protease Inhibitors

Stefins or type 1 cystatins belong to a large, evolutionarily conserved protein superfamily, the members of which inhibit the papain-like cysteine proteases. Three mouse stefins (1, 2, and 3) were isolated on the basis of their relatively increased expression in motheaten viable compared to normal congenic mouse bone marrow cells. The open reading frames of the stefin cDNAs encoded proteins of approximately 11.5 kilodaltons that show between 50% and 92% identity to sequences of stefins isolated from various other species. Data from Southern analysis suggested that the murine stefin gene family encompasses at least 6 and possibly 10 to 20 members, all of which appear to be clustered in the genome. Analysis of IB mice indicates that the genes encoding the three mouse stefins all map to mouse chromosome 16, a localization that is consistent with the recent assignment of the human stefin A gene to a region of conserved homology between human chromosome 3q and the proximal region of mouse chromosome 16.

Molecular Cloning, Chromosomal Location, and Tissue-Specific Expression of the Murine Cathepsin G Gene

A cluster of genes encoding cathepsin G (CG) and two other CG-like hematopoietic serine proteases, CGL-I and CGL-2, was previously mapped to human chromosome 14. Using a human CG (hCG) cDNA as the probe, a novel, related murine hematopoietic serine protease gene was isolated. This murine gene spans ~2.5 kilobases of genomic DNA, is organized into five exons and four introns, and bears a high degree of homology to hCG at both nucleic acid (73%) and deduced amino acid (66%) levels. The predicted cDNA contains an open reading frame of 783 nucleotides that encodes a nascent protein of 261 amino acids. Processing of a putative signal (pre) peptide of 18 residues and an activation (pro) dipeptide would generate a mature enzyme of ~27 kilodaltons that has an estimated pI of 12.0. Conserved residues at His44, Asp88, and Ser181 form the characteristic catalytic triad of the serine protease superfamily. The gene was tightly

linked to the *CTLA-1* locus on murine chromosome 14, where the serine protease genes *mCCP1-4* are clustered. Expression of this gene was detected only in the bone marrow and was restricted to a small population of early myeloid cells.

Gtx: A Novel Murine Homeobox-Containing Gene, Expressed Specifically in Glial Cells of the Brain and Germ Cells of Testis, Has a Transcriptional Repressor Activity in Vitro for a Serum-Inducible Promoter

Although it is likely that a highly complex network of transcription factors acts in concert during mammalian brain development, relatively few such genes have been characterized to date. A novel murine homeobox gene, denoted *Gtx*, was isolated. In adult animals, *Gtx* is specifically expressed within glial cells of the central nervous system, including the forebrain, and in germ cells of the testis. *Gtx* was found to reside on mouse chromosome 7 and did not cosegregate with any previously mapped homeobox gene. The amino acid sequence of the predicted protein encoded by *Gtx* is highly divergent from that of any other known homeobox genes. The *Gtx* homeodomain contains unique residues at positions predicted to contact DNA bases. GTX did not bind to known target sites for other homeobox gene products in vitro but bound with high affinity to the MEF-2 motif, a binding site for the serum response factor-related proteins. GTX efficiently competed with RSRF to bind the MEF-2 element in vitro. Although the true biological role of *Gtx* is not known, these results suggest that *Gtx* is a novel cell-type-specific homeobox gene that has the potential to act as a transcriptional repressor for a subset of serum-inducible genes.

A Novel Class of Murine POU Gene Predominantly Expressed in the Central Nervous System

A mouse gene (referred to as *Emb*) encoding a novel class of POU domain proteins was identified. The Emb POU domain shares only 40% to 50% homology with any other POU proteins. Nonetheless, the Emb POU domain can bind to the octamer sequence like other POU domain proteins. *Emb* is a single-copy gene that was localized to the distal region of mouse chromosome 15. *Emb* is expressed in the embryo throughout postimplantation stages, where the most prominent expression was seen in the developing central nervous system. In the adult, *Emb* is highly expressed in brain, whereas weaker expression was detected in other organs such as testis, skeletal muscle, and kidney. The expression in adult brain was most evident in neurons of hypocampus formation. Two types of *Emb* mRNA are expressed in brain: one type encodes a protein of 301 amino acid residues, whereas the other codes for a protein with two extra amino acid residues added at the N terminus of the POU domain. These two mRNA species are generated by alternative splicing by utilizing an unusual splice acceptor site: CCTCCCTCTG/. *Emb* mRNA expressed in testis, on the other hand, encodes a smaller protein lacking most of the N-terminal region.

Cloning and Structure of the Human Corticotrophin Releasing Factor-Binding Protein Gene

The human corticotrophin-releasing factor-binding protein gene (*CRHBP*) was cloned and mapped to the distal region of chromosome 13 and loci 5q in the mouse and human genomes, respectively. The gene consists of 7 exons and 6 introns. The mature protein has 10 cysteines and 5 tandem disulfide bridges, 4 of which are contained within exons 3, 5, 6, and 7. One bridge is shared by exons 3 and 4. The signal peptide and the first 3 amino acids of the mature protein are encoded for by an extreme 5′ exon. Primer extension analyses revealed that the transcriptional initiation site is located 32 base pairs downstream from a consensus TATA box. The promoter sequence contained a number of putative promoter elements, including an AP-1 site; three estrogen receptor (ER) half sites; sites for the immunoglobulin enhancer binding proteins, NF-κB and INF-1; and the liver-specific enhancer binding proteins, LFA1 and LFB1.

S-laminin: Mapping to Mouse Chromosome 9 and Expression in the Linked Mutants *tippy* and *ducky*

S-laminin, a homolog of the laminin B1 chain, is present in a subset of basal laminae, including those of the skeletal neuromuscular junction and the renal glomerulus. We showed that the distribution and apparent size of murine S-laminin is similar to those documented previously for rat and human. Using IB analysis, we mapped

the *S-laminin* (*Lams*) gene to mouse chromosome 9. Thus, it is unlinked to genes for the laminin A, B1, and B2 chains. Finally, because the *Lams* gene maps near two mutations that affect neuromuscular function, ducky (*du*) and tippy (*tip*), we assayed S-laminin by Southern blotting, immunoblotting, and immunohistochemistry in these mutants. No abnormality of the S-laminin gene or protein was detectable in either mutant.

Reversal of Left-Right Asymmetry: A Situs Inversus Mutation

A recessive mutation was identified in a family of transgenic mice that resulted in a reversal of left-right polarity (situs inversus) in 100% of the homozygous transgenic mice tested. Sequences that flanked the transgenic integration site were cloned and mapped to mouse chromosome 4, between the *Tsha* and *Hxb* loci. During early embryonic development, the direction of postimplantation turning, one of the earliest manifestations of left-right asymmetry, was reversed in homozygous transgenic embryos. This insertional mutation appears to identify a gene that controls embryonic turning and visceral left-right polarity.

Tissue-Specific Expression and Chromosomal Localization of the α Subunit of Mouse Meprin A

Meprins, membrane-bound oligomeric metalloendopeptidases, contain α and/or β subunits. Their activities have been found in the mouse and rat kidney. The cloned cDNA for the mouse α subunit of meprin A was used to survey mRNA expression in the kidney of different mouse strains and in various tissues of mice and rats. A single message of 3.6 kilobases was found in the kidney of random bred (ICR) and inbred mice (C57BL/6, DBA/2) that contain high meprin A activity and in Sprague-Dawley rat kidney. The α subunit message was undetectable in the kidney of C3H/He and CBA mice, inbred strains that do not express meprin A activity. Therefore, meprin A activity in the kidney of mouse strains correlates with the amount of α subunit mRNA present. The 3.6-kilobase mRNA meprin α subunit message was also detected in the small intestine of the rat but not in mice. No message was detected in brain, heart, skeletal muscle, liver, lung, or spleen of mice or rats. PCR amplification or Southern blot analysis of genomic DNA revealed that the gene for the α subunit is present in all mouse strains as well as in human, monkey, rat, mouse, dog, cow, rabbit, and chicken, but it was not detected in yeast. There is one gene copy present in the mouse genome. The gene was localized to mouse chromosome 17 centromeric to the major histocompatibility complex (H-2) by IB analysis. The localization of this allele to *Mep-1*, the gene previously found to regulate the expression of meprin A activity in mice, supports the proposal that *Mep-1* is the structural gene for the α subunit.

Localization of a Novel Natural Killer Triggering Receptor Locus to Human Chromosome 3p23-p21 and Mouse Chromosome 9

A novel gene (*NKTR*) that is involved in the recognition of tumor cells by large granular lymphocytes was assigned to the short arm of human chromosome 3 in the region 3p23-p21 by somatic cell hybrid analysis. IB analysis revealed that the murine homolog maps to the distal end of mouse chromosome 9 and is closely linked to the locus coding for cholecystokinin. This region of mouse chromosome 9 shares a region of homology with human chromosome 3p. Thus, the placement of *NKTR* in these regions confirms and extends the relationship between these human and mouse chromosomes.

Molecular Cloning of a Diverged Homeobox Gene That Is Rapidly Downregulated During the G_0/G_1 Transition in Vascular Smooth Muscle Cells

Adult vascular smooth muscle cells dedifferentiate and reenter the cell cycle in response to growth factor stimulation. A diverged homeobox gene, *Gax*, whose expression is largely confined to the cardiovascular tissues of the adult was isolated. In quiescent adult rat vascular smooth muscle cells, *Gax* mRNA levels were downregulated as much as 15-fold within 2 hours when these cells were induced to proliferate with platelet-derived growth factor (PDGF) or serum growth factors. This reduction in *Gax* mRNA was transient, with levels beginning to rise between 8 and 24 hours after mitogen stimulation and returning to near normal by 24 to 48 hours. The *Gax* downregulation was

dose dependent and could be correlated with the mitogen's ability to stimulate DNA synthesis. PDGF-AA, a weak mitogen for rat vascular smooth muscle cells, did not affect *Gax* transcript levels, while PDGF-AB and -BB, potent mitogens for these cells, were nearly as effective as fetal bovine serum. The removal of serum from growing cells induced *Gax* expression fivefold within 24 hours. These data suggested that *Gax* is likely to have a regulatory function in the G_0/G_1 transition of the cell cycle in vascular smooth muscle cells.

TrkC, a Receptor for Neurotrophin-3, Is Widely Expressed in the Developing Nervous System and in Non-Neuronal Tissues

The Trk family of tyrosine kinases encodes receptors for nerve growth factor-related neurotrophins. The developmental expression of *trkC*, which encodes a receptor for neurotrophin-3 (NT-3), was studied. Like the related genes *trk* and *trk*B, *trk*C was expressed primarily in neural lineages, although the pattern is complex and includes non-neuronal cells. Direct comparison with *trk* and *trk*B developmental expression patterns permitted the following observations. (1) *trk*C is expressed in novel neural tissues where other *trk* genes are silent. (2) Some tissues appear to coexpress *trk*B and *trk*C receptors in the embryo and in the adult. (3) *trk*C expression can be detected in the gastrulating embryo. These data provide insights into the role of Trk family receptors and nerve growth factor-related neurotrophins during development and suggest that, in addition to regulating neuronal survival and differentiation, the neurotrophin/Trk receptor system may have broader physiological effects. IB analysis was used to map the location of each of the *trk* genes on mouse chromosomes. Alignment with available chromosomal maps identified possible linkage between the *trk* genes and known neurological mutations.

Structural Comparison and Chromosomal Localization of the Human and Mouse IL-13 Genes

The genomic structure of the recently described cytokine IL-13 was determined for both human and mouse genes. The human IL-13 gene (*IL13*) occurs as a single copy in the haploid genome and maps to human chromosome 5. A 4.3-kilobase DNA fragment of the mouse IL-13 gene (*Il13*) was sequenced and found to occur as a single-copy gene, mapping to mouse chromosome 11. Intrachromosomal mapping studies revealed that both human and mouse genes contain four exons and three introns and show a high degree of sequence identity throughout their length. Potential recognition sequences for transcription factors are present in the 5′ flanking region and are conserved between both genes. Both genes map to chromosomal locations adjacent to genes encoding other cytokines, including IL-3, GM-CSF, IL-5, and IL-4, suggesting that IL-13 is another member of this cytokine gene family that may have arisen by gene duplication.

CD30 Antigen, a Marker for Hodgkin's Lymphoma, Is a Receptor Whose Ligand Defines an Emerging Family of Cytokines with Homology to TNF

CD30 is a surface marker for neoplastic cells of Hodgkin's lymphoma and shows sequence homology to members of the tumor necrosis factor (TNF) receptor superfamily. Using a chimeric probe consisting of the extracellular domain of CD30 fused to truncated immunoglobulin heavy chains, a CD30 cDNA cognate was cloned from the murine T-cell clone 7B9 expression library. The encoded protein was a 239-amino acid type II membrane protein whose C-terminal domain showed significant homology to TNF, TNFβ, and CD40L. Cross-hybridization to an induced peripheral blood T-cell cDNA library yielded the human homolog, which was 72% identical at the amino acid level. The recombinant human ligand enhanced the proliferation of CD3-activated T cells yet induced differential responses, including cell death, in several CD30+ lymphoma-derived clones. The human and murine genes mapped to 9q33 and the proximal region of chromosome 4, respectively.

GHRH Receptor of *Little* Mice Contains a Missense Mutation in the Extracellular Domain That Disrupts Receptor Function

The growth hormone-releasing hormone receptor (GHRHR) is a member of the family of G protein-coupled receptors that is expressed on pituitary cells and mediates the actions of GHRH in stimulating growth hormone (GH)

synthesis and secretion. The Ghrh receptor gene was found to be located in the middle of mouse chromosome 6 in the same region as the *little* mutation. Mice homozygous for this mutation have reduced GH secretion and a dwarf phenotype. A missense mutation was identified in the extracellular domain of the *little* GHRHR that disrupts receptor function, suggesting that the growth deficit in these mice results from a defect in the GHRHR. Similar alterations in GHRHR might explain some isolated GH deficiencies in humans.

Species-Specific Induction of Cytochrome P4504A RNAs: PCR Cloning of Partial Guinea Pig, Human, and Mouse *CYP4A* cDNAs

PCR amplification was used to demonstrate the presence of a conserved region and to clone novel members of the cytochrome P4504A gene family from guinea pig, human, and mouse cDNAs; this strategy was based on the sequences at positions 925 to 959 and the heme-binding domain (1381 to 1410) of the rat *Cyp4a1* gene. Murine *Cyp4a* clones were highly homologous to members of the rat gene family, but *Cyp4a* clones from human and guinea pig were equally similar to the rat/mouse genes, suggesting that the rat/mouse line had undergone gene duplication events after divergence from human and guinea pig lines. The mouse *Cyp4a12* clone was localized to chromosome 4 by IB mapping, in a region of synteny with human chromosome 1. The assignment of the human *CYP4A11* gene to chromosome 1 was confirmed by somatic cell hybridization. An RNase protection assay was shown to discriminate between the murine *Cyp4a10* and *Cyp4a12* cDNAs. Treatment of mice with the potent peroxisome proliferator, methylclofenapate, induced *Cyp4a10* RNA in liver, and to a lesser extent in kidney; there was no sex difference in response. *Cyp4a12* was present at high levels in male control liver and kidney samples and was not induced by treatment with methylclofenapate. *Cyp4a12* was present at low levels in control female liver and kidney RNA and was highly induced in both organs by methyclofenapate. Guinea pigs were exposed to methylclofenapate (50 mg/kg), but there was no significant induction of the guinea pig *Cyp4a13* RNA. These findings are consistent with a species difference in response to peroxisome proliferators between rat/mouse and guinea pig.

Unp, a Mouse Gene Related to the *tre* Oncogene

A novel gene designated *Unp* was identified. *Unp* cDNAs contained a large open reading frame that could encode a protein of 89 kilodaltons. The predicted protein contains a putative nuclear localization signal, as well as consensus sequences for binding to the retinoblastoma gene product. The latter elements are contained within a region having strong similarity to the human *tre* oncogene. The *Unp* gene was localized to mouse chromosome 9 in a region of homology with human chromosome 3p. This region has been implicated in a number of human malignancies.

Organization of the Murine *Cd22* Locus

Murine CD22 (mCD22) is a B-cell–associated adhesion protein with seven extracellular Ig-like domains that has 62% amino acid identity to its human homolog. Southern analysis of genomic DNA isolated from tissues and cell lines from several mouse strains with mCD22 cDNA demonstrated that the *Cd22* locus encoding mCD22 is a single-copy gene of ≤30 kilobases. Digestion of genomic DNA preparations with four restriction endonucleases revealed the presence of RFLPs in BALB/c, C57BL/6, and C3H strains versus DBA/2J, NZBV, and NZC strains, suggesting the presence of two or more *Cd22* alleles. Using an mCD22 cDNA clone derived from the BALB/c strain, genomic clones from a DBA/2J genomic library were isolated that contained all the exons necessary to encode the full-length mCD22 cDNA. Fifteen exons, including exon 3 that encodes the translation start codon, were identified. Each extracellular Ig-like domain of mCD22 was encoded by a single exon. A comparison between the nucleotide sequences of the BALB/c CD22 cDNA and the exons of the DBA/2J CD22 genomic clones revealed an 18-nucleotide deletion in exon 4 (encoding the most distal Ig-like domain 1 of mCD22) of the DBA/2J genomic sequence in addition to a number of substitutions, insertions, and deletions in other exons. These nucleotide differences were also present in a cDNA clone isolated from total RNA of LPS-activated DBA/2J splenocytes by RT-PCR. The *Cd22* locus was mapped to the proximal region of chromosome 7, a region syntenic to human chromosome 19q, close to the previously reported loci *Lyb-8* and *Mag* (a homolog of *Cd22*). An antibody (CY34) against the Lyb-8.2 B-cell marker reacted with a

transfected cell line expressing the full-length mCD22 cDNA, thus demonstrating that *Lyb-8* and *Cd22* loci are identical.

Mutation in the Unique Region of Bruton's Tyrosine Kinase in Immunodeficient XID Mice

The cytoplasmic tyrosine kinase, Bruton's tyrosine kinase (Btk, formerly bpk or atk), is crucial for B-cell development. Loss of kinase activity results in the human immunodeficiency X-linked agammaglobulinemia, characterized by a failure to produce B cells. In the murine X-linked immunodeficiency (XID), B cells are present but respond abnormally to activating signals. The Btk gene, *btk*, was mapped to the *xid* region of the mouse X chromosome by IB analysis. A single conserved residue within the N-terminal unique region of Btk was mutated in XID mice. This change in the *xid* gene probably results in interference with normal B-cell signaling mediated by Btk protein interactions.

Molecular Embryology Section

Luis F. Parada

Xi He
Marie Mazzulla
Mary Ellen Palko
James Pickel
Susan W. Reid
Dan Soppet
Kenji Suzuki
Yoshio Taniyama
Lino Tessarollo
Pantelis Tsoulfas
Alfons Valera
Kristine Vogel

Our objective is to gain a molecular understanding of the regulatory pathways that control the mammalian embryogenesis. We use the mouse as a primary model system for our studies. However, in the past year, we have expanded the scope of our studies to using frogs and birds in situations where these organisms have provided better suited systems for experimentation.

Through the investigation of embryonic expression patterns for several oncogene families, we have identified proto-oncogenes expressed primarily during embryogenesis in spatio-temporally restricted profiles that suggest important regulatory roles.

Trk Genes

Over the last five years, we have reported on our continuing studies of the *trk* gene family of receptor tyrosine kinases (RTKs; Figure 1). A salient feature of this gene family is the predominant expression of *trk* genes in the nervous system. We have previously investigated the expression of two members (*trk* and *trk*B) during embryogenesis and in the adult. The *trk* gene is expressed primarily in neural crest-derived sensory neurons from the onset of neurogenesis, throughout development, and in the adult nervous system. This restricted expression pattern provided

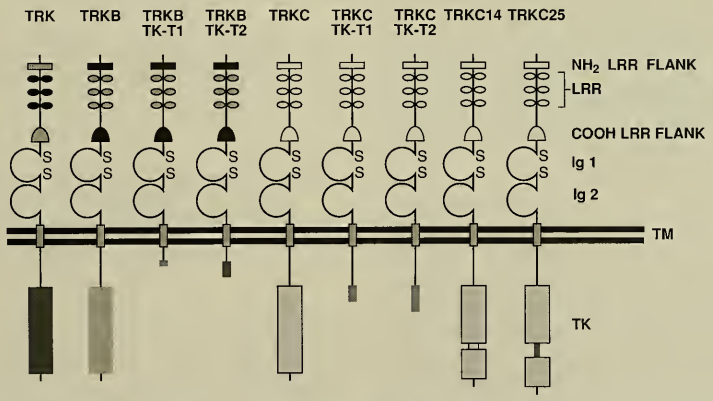


FIGURE 1. Schematic representation of different members and related isoforms of the *trk* gene family. The N-terminal sequences of the proteins contain the flank–leucine-rich repeat–flank structural motif followed by two tandem repeats of immunoglobulin (Ig)-like domains. TM, transmembrane domain; TK, tyrosine kinase domain.

important initial clues regarding the identity of its ligand, the neurotrophin nerve growth factor (NGF). The *trk*B gene exhibits a broader expression pattern than that of *trk*, although *trk*B is also predominantly expressed throughout the neural network (central and peripheral nervous systems) from early neural development and in the mature nervous system. In addition to a complex expression profile, the structure of the *trk*B gene is also complex. Multiple transcripts and at least two polypeptide isoforms, which arise from alternatively spliced transcripts, are produced by this locus. Our past work has contributed to the identification of the TrkB receptor as a promiscuous receptor that is capable of interacting with several of the NGF-related neurotrophins including brain-derived neurotrophic factor (BDNF) and neurotrophins 3 and 4/5 (NT-3 and NT-4/5). More recently, we and others have studied a third *trk* family member, *trk*C, which is also broadly expressed in the nervous system and whose ligand recognition is limited to NT-3.

In collaboration with Neal Copeland and Nancy Jenkins, Molecular Genetics of Oncogenesis and Molecular Genetics of Development Sections, we have mapped the location of the three *trk* genes in mouse chromosomes through *Mus musculus/Mus spretus* backcross analysis. The *trk* gene maps to mouse chromosome 4, *trk*B to mouse chromosome 13, and *trk*C to mouse chromosome 7. Each of the genes maps to regions that contain genes responsible for neurological development in the mouse. The *trk* gene maps near the spastic (*spa*) locus, *trk*B near the Purkinje cell degeneration (*pcd*) locus, and *trk*C near a region that includes genes identified by mutations that are associated with a series of inner ear and hearing defects. Whether the mutations at the *trk* loci are causes for any of these defects should be determined in the near future.

Trk Family Proteins

To examine the products of the various *trk* loci, and to verify the existence of the proteins predicted by cDNA isolation and sequencing, we have generated receptor-specific antisera, as well as antisera that recognize multiple Trk receptors. These antibodies have permitted the demonstration of direct receptor/ligand interaction between the various Trk family receptors and the neurotrophins. Figure 2 shows a Western analysis assaying Trk proteins

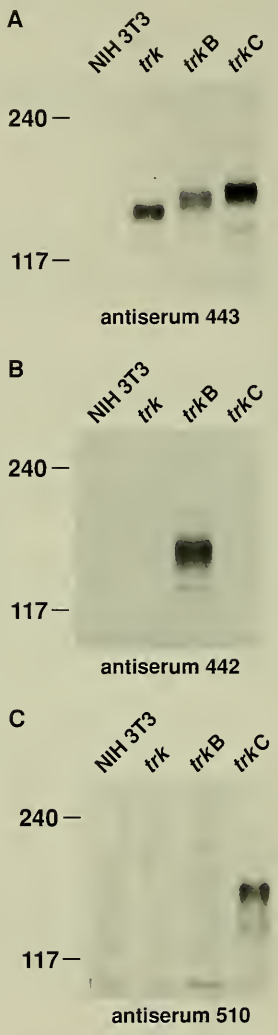


FIGURE 2. Untreated NIH/3T3 cells or NIH/3T3 cells stably transfected with *trk*, *trk*B, or *trk*C were subjected to Western blot analysis with rabbit polyclonal antiserum 443, which recognizes Trk, TrkB, and TrkC (A); TrkB-specific antiserum 442 (B); or TrkC-specific antiserum 510 (C). The apparent molecular weights (X1000) are indicated on the left.

from NIH/3T3 cell lines transfected with cDNAs encoding either *trk*, *trk*B, or *trk*C. Antiserum 443, which recognizes the conserved C terminus of *trk* family-encoded proteins, readily detects gp140$^{proto trk}$, gp145^{trkB}, and gp150^{trkC} (Figure 2A). The antisera 442 and 510 have been generated from peptides corresponding to the less well conserved extracellular domains of gp145^{trkB} and gp150^{trkC}, respectively. Antiserum 442 specifically recognizes gp145^{trkB} (Figure 2B), while antiserum 510 recognizes gp150^{trkC} (Figure 2C) and gp150^{trkC14} (data not shown). Similar analysis of tissue extracts has demonstrated the presence of gp140$^{proto trk}$ in embryonic dorsal root ganglia (DRG); gp145^{trkB}, gp110$^{trkBTK-}$, gp150^{trkC}, and gp120$^{trkCTK-}$ in adult brain (Figure 3); and gp150^{trkC14} in adult brain (data not shown). Although *trk*B and *trk*C transcripts are present in many tissues, we do not detect these proteins in tissues outside the nervous system. This observation may reflect the greater sensitivity of transcript detection methods relative to the sensitivity of our antisera.

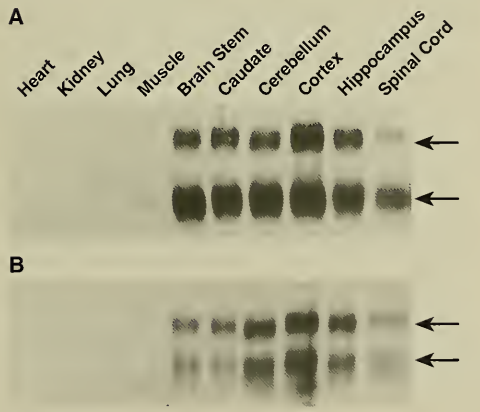


FIGURE 3. Expression of TrkB and TrkC in mouse tissues. Western blot analysis of mouse tissues with TrkB-specific antiserum 442 (A) or TrkC-specific antiserum 510 (B). The lanes contain equivalent amounts of protein obtained from the tissues indicated at the top of the figure. The position of the 145- to 150-kilodalton full-length protein (upper arrow in each panel) is indicated on the right side of the figure. The lower band (arrow) indicates truncated TrkB and TrkC receptors.

TrkB

While in fibroblasts, the TrkB receptor can bind multiple ligands with similar affinities and, in consequence, becomes phosphorylated on tyrosine residues (Figure 4); the biological significance of these interactions is not yet entirely clear. This apparent in vitro receptor promiscuity must be reconciled with the in vivo observations that specific neurons most often respond to a single neurotrophin rather than multiple neurotrophins.

In an effort to better understand this interaction, we have expressed TrkB receptors in PC12 cells. These cells undergo neuronal differentiation in response to NGF (via Trk receptors) and BDNF and NT4/5 (via TrkB receptors; Figure 5). However, the TrkB receptor does not mediate a differentiation response to NT-3 in PC12 cells, even though when expressed in fibroblasts or in frog oocytes, the receptor is capable of binding the neurotrophin and of transducing proliferative or maturation signals to the nucleus. We do not currently understand the molecular basis for TrkB ligand discrimination in this PC12 assay, but our continuing and future efforts are focused on elucidating this underlying mechanism.

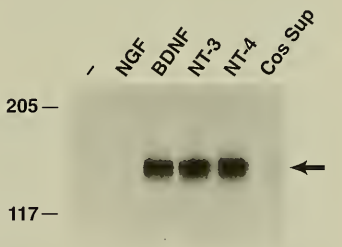
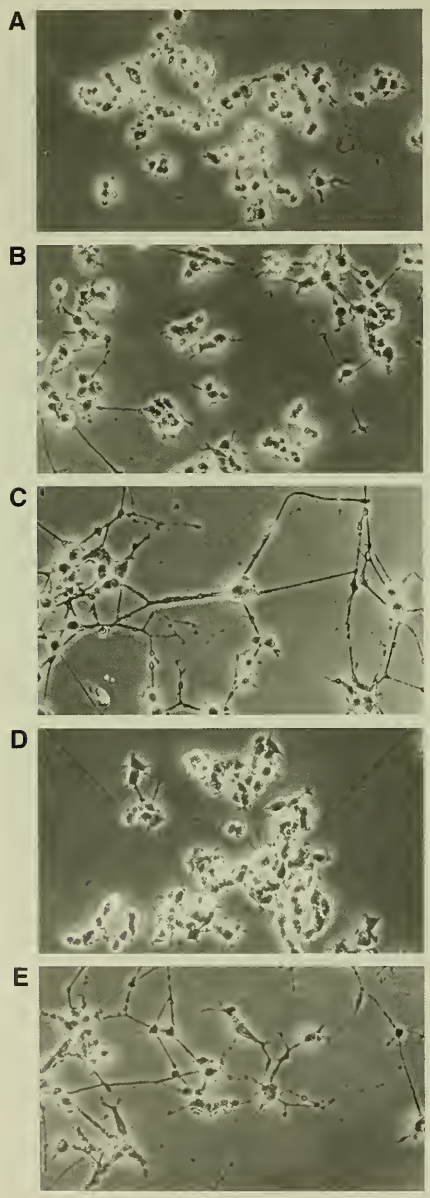


FIGURE 4. TrkB is phosphorylated on tyrosine in response to BDNF, NT-3, and NT-4. NIH/3T3 cells expressing *trk*B were exposed to neurotrophins for 5 minutes at 37°C. Extracts from these cells were prepared in the presence of 1 mM orthovanadate, immunoprecipitated with TrkB-specific antisera, and resolved by SDS-polyacrylamide gel electrophoresis. Tyrosine phosphorylation was determined by Western blot analysis using the antiphosphotyrosine monoclonal antibody 4G10. The lanes include untreated *trk*B-expressing cells (–) as well as cells treated with NGF, BDNF, NT-3, NT-4 expressed in Cos cells, or Cos cell supernatant (Cos Sup). The apparent molecular weights of standards (X1000) are indicated on the left and the position of the *trk*B-encoded protein is indicated by the arrow on the right.



TrkC

Last year, we reported the cloning of a third *trk*-related gene, *trk*C, and we showed that two isoforms, including one that contains a 14-amino acid insertion within the catalytic domain (TrkC14), function as receptors for the neurotrophin NT-3 when expressed in NIH/3T3 fibroblasts. As shown in Figure 1, there exists at least a third full-length TrkC receptor, which harbors a 25-amino acid insert in its catalytic domain (TrkC25).

We have studied *trk*C gene expression and compared it to that of *trk* and *trk*B. The earliest expression of *trk*C was observed in the gastrulating embryo, coordinate with the temporal expression pattern of its ligand NT-3, the most abundant neurotrophin during early development. This is the earliest example of *trk* gene activity reported to date, raising the possibility that the *trk*C RTK mediates important functions in early neuroepithelium formation. We used RNase protection assays, employing a β-actin internal control, to assess the expression of *trk*C in development. Although a faint protected band is first observed on embryonic day 7.5, *trk*C gene expression is clearly detected at embryonic day 8.5, coinciding with the timing of neural tube formation (Figure 6). Relative to the internal β-actin control, a substantial increase in *trk*C mRNA expression occurred around embryonic day 10.5.

We next performed RNA in situ hybridization on sections from postimplantation embryos (day 7.5 through day 17.5) and in neonate and adult central nervous system (CNS). The results obtained from our in situ experiments on *trk*C expression are summarized in Table 1. The expression of this gene is highly complex. We first observed *trk*C hybridization as a weak signal in embryonic day 7.5 egg cylinders with a pattern indicating expression in the early neuroectoderm (not shown). Expression of *trk*C is evident throughout the neuroepithelium of day 9.5 embryos. At these early stages, we also observed *trk*C transcripts in

FIGURE 5. PC12 cells expressing *trk*B do not respond to NT-3. *Trk*B-expressing PC12 cells were plated on collagen-coated tissue culture dishes. The cells were treated with PC12 medium alone (A), PC12 medium supplemented with 50 ng/ml NGF (B), 10 ng/ml BDNF (C), 20 ng/ml NT-3 (D), or 10 ng/ml NT-5 (E). Representative fields were photographed by phase-contrast microscopy 5 days after treatment.

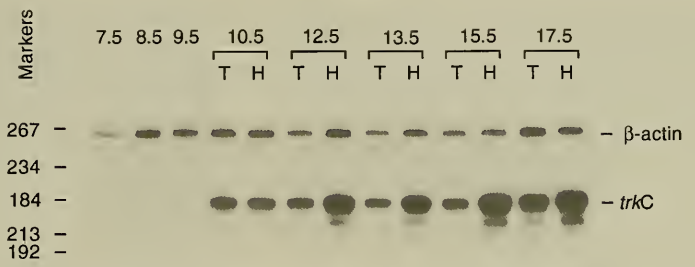


FIGURE 6. RNase protection analysis of *trk*C expression in mouse embryos from 7.5 to 17.5 days of gestation. Approximately 10 μg of total RNA from embryonic trunks (T) and heads (H) were analyzed using a ^{32}P-labeled *trk*C-specific probe. A β-actin probe was used in the same hybridization mixture with *trk*C as an internal control for RNA quantitation. The positions of coelectrophoresed DNA size markers (in base pairs) are indicated at the left.

TABLE I. Summary of *trkC* expression

	9.5E	11.5E	13.5E	15.5E	17.5E
(A) During mouse development					
CNS	Prosencephalon Mesencephalon Rhomboencephalon	Telencephalon Diencephalon Mesencephalon Rhomboencephalon	See panel B	See panel B	See panel B
	Neural tube	Neural tube	Cerebellum Spinal cord	Cerebellum Spinal cord	Cerebellum Spinal cord
PNS	Neural crest cells	DRG V trigeminal VII-VIII facio-acoustic complex IX superior and inferior glossopharyngeal X superior ganglion	DRG V trigeminal VIII vestibulococlear ganglia Intestine coeliac ganglion	DRG V trigeminal VIII vestibulococlear ganglia Intestinal ganglia Sympathetic	DRG V trigeminal VIII vestibulococlear ganglia Enteric ganglia Sympathetic
Others	Aortic arch arteries	Dorsal aorta Urogenital mesenchyme	Aorta Arteries Urogenital mesenchyme	Aorta Lung arteries Adrenal Submandibular gland Cervical brown adipose Foot pad Whisker pad	Aorta Lung arteries Adrenal Submandibular gland Cervical brown adipos Foot pad Whisker pad

	13.5E	15.5E	17.5E	10-day	6-week
(B) In the CNS					
	Neocortex Basal telencephalon Hippocampal neo Striatum Diencephalon Pons Medulla Cerebellum Facial nucleus	Neocortex Cortical plate Intermediate layer Striatum Pons Medulla Cerebellum External germinal layer Purkinje cell layer	Neocortex Cortical plate Intermediate layer Striatum Pons Medulla Cerebellum External germinal layer Purkinje cell layer	Cortex Caudate putamen Fornix Thalamic nuclei Hippocampus Medulla Olfactory nucleus Cerebellum Granule cell layer Purkinje cell layer	Cortex Hypothalamus Thalamic nuclei Hippocampus Olfactory nucleus Cerebellum Granule cell layer Purkinje cell layer

regions where neural crest cells localize, including the forming DRG and regions adjacent to the neural tube and the dorsal aorta, locations identified as migratory routes used by cells of neural crest origin (see below).

During organogenesis, *trk*C expression is exhibited throughout the CNS (in the telencephalon, diencephalon, mesencephalon, rhombencephalon, and neural tube). Comparison with an adjacent *trk*B-hybridized section (not shown) indicates overlapping expression of *trk*B and *trk*C in the mantle (postmitotic) layer of the spinal cord, whereas the ependymal (mitotic) layer expresses only *trk*B. The *trk*B and *trk*C genes are coexpressed in the branchial arch arteries; however, more caudal, *trk*B transcripts are present in the mesonephric region, while *trk*C is expressed in the midgut and genital ridge-associated mesenchyme.

At 13.5 days of development, the *trk*C gene is expressed throughout the neural network including neural tissues in which *trk* and *trk*B gene activity has not been previously detected. One striking example is the expression of *trk*C in neural crest-derived cells of the developing enteric nervous system (Figure 7). The ganglia of the enteric nervous system are formed from migratory neural crest cells originating in the vagal crest region. These cells hybridized with *trk*C probes as exemplified by hybridization of embryonic day 13.5 stomach (Figure 7A), where *trk*C-expressing migratory neural crest cells are intercalated within the mesenchyme. Several days later, *trk*C transcripts could be localized to the forming ganglia that appear throughout the muscle and submucosal layers of the intestine (Figure 7B). We did not observe *trk*C expression in the epithelium (Figure 7). Thus, *trk*C expression is found in diverse peripheral nervous system (PNS) structures, some of which do not appear to express other Trk family receptors.

Several insights can be drawn from a direct comparison of *trk*, *trk*B, and *trk*C expression at later stages in development. Figure 8 shows in situ hybridization of sections from a single embryonic day 17.5 fetus that were hybridized with probes to the three related genes. Expression of *trk* is distinctive in its tight regulation: *trk* RNA is primarily limited to trigeminal (V), superior cervical (C), sympathetic (S), and DRG (arrows). In comparison, *trk*B is also expressed in these ganglia and, at high levels, in the CNS and additional structures, including the whisker pad (W), tongue (T), and tooth buds (white triangle; Figure 8B). This complex expression pattern is consistent throughout mid-gestation (days 14 to 18). Expression of the *trk*C gene mirrors that of *trk*B in many regions of the embryo (Figure 8B and 8C). Like the *trk*B gene, *trk*C is expressed throughout the nervous system, including many of the same cranial and spinal ganglia (V, VII, VIII, X, DRG). The *trk*C and *trk*B genes are also apparently coexpressed in several non-neural structures, including the tongue, the whisker pad (vibrissae) mesenchyme, and forming teeth (Figure 8B and 8C). In the teeth, *trk*B is expressed in the outer dermal papilla, while *trk*C transcripts appear confined to the core of the dermal papilla (compare Figure 8, panels B and C).

As shown in Figure 8, *trk*C-encoding transcripts are abundant throughout the embryo, including sites that are not necessarily consistent with *trk*C expression in neural cells (see Table 1): for example, *trk*C transcripts are present in vibrissae and in dental papillae (Figure 8). We also found *trk*C transcripts in cells of the submandibular gland (Figure 9A and 9B), and in the mesenchyme surrounding mesonephric and urogenital ducts (Figure 9C and 9D). High levels of *trk*C transcripts could also be localized to the brown adipose tissue dorsal to the cervical spinal cord (Figure 9E and 9F) and the cortex of the metanephros and adrenal gland (Figure 9G and 9H). Finally, abundant *trk*C expression was found along the subendothelial mesenchyme of arteries throughout the developing embryo (Figure 9I and 9J).

We have previously demonstrated that NT-3 specifically binds and induces tyrosine autophosphorylation of TrkC and TrkC14 in fibroblasts. However, these two receptor isoforms gave different results when expressed in PC12 cells (Figure 10). The appearance of neurites in response to added NT-3 was observed only in cultures that were transfected with the *trk*C plasmid, not with the *trk*C14 plasmid. We have now extended this analysis to include a third TrkC isoform (TrkC25) (see Figure 1). We find that, like TrkC14, TrkC25 is inactive in the PC12 neurite outgrowth assay, even though the receptor can be autophosphorylated in response to NT-3 addition.

We assayed stably transfected *trk*C, *trk*C14, and *trk*C25 cell lines for neurite outgrowth (Figure 10). Stimulation of these cells with NGF indicated that the neurite outgrowth pathway via the Trk receptor was intact in all clones tested. However, while NT-3 elicited neurite outgrowth in *trk*C-expressing PC12 cells over a broad range of ligand

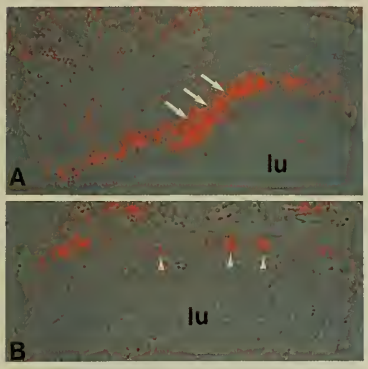


FIGURE 7. Expression of *trk*C in the developing gut. Bright-dark-field view of sections through embryonic day 13.5 stomach (A) and day 17.5 intestine (B); lu indicates the stomach (A) and intestine (B) lumen; arrows in panel A point at migratory neural crest cells; arrowheads in panel B indicate the forming myenteric plexus.

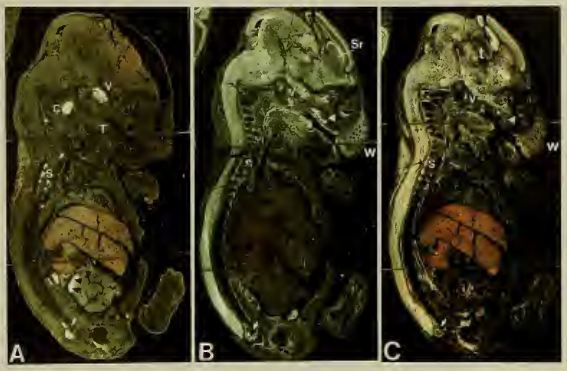


FIGURE 8. Comparative embryonic expression of *trk* family genes. Dark-field views of serial adjacent sagittal sections of a 17.5 day mouse embryo hybridized with a *trk* (A), *trk*B (B) or *trk*C (C)-specific probe. White triangles indicate the tooth papilla; arrows, DRG; Sr, striatum; t, thalamus; T, tongue; V, fifth cranial ganglion; C, superior cervical ganglion; S, sympathetic ganglia; W, whisker pad; i, intestine. The black triangles in panel A point to artifactual staining.

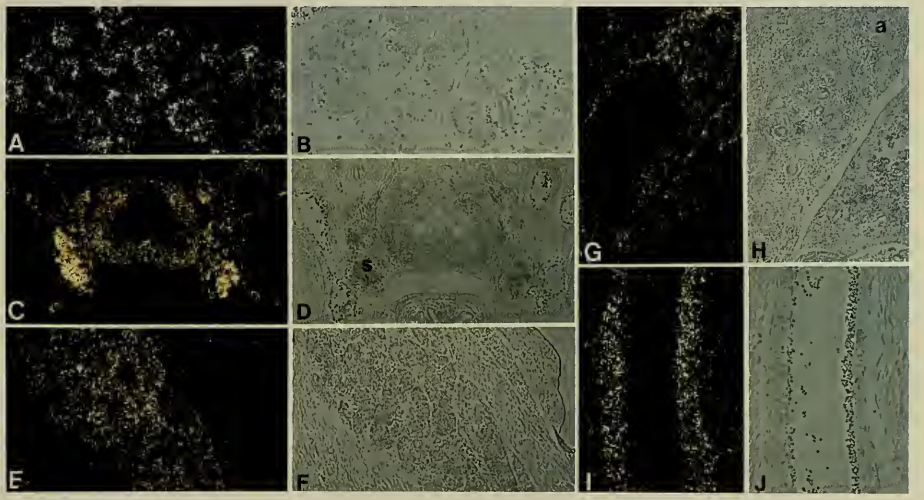


FIGURE 9. Expression of *trk*C in non-neural day 15.5 embryonic tissues. Shown are dark-field (A) and light-field (B) views of the submandibular salivary gland; dark-field (C) and bright-field (D) views of a section through the genital region (s, sympathetic ganglia); dark-field (E) and light-field (F) views of cervical brown adipose tissue; dark-field (G) and light-field (H) views of a section through the kidney and the adrenal gland (a); and dark-field (I) and light-field (J) views of a sagittal section through the dorsal aorta. Dark-field optics were obtained using a Schott fiber optic condenser attached to a Zeiss axioplot microscope.

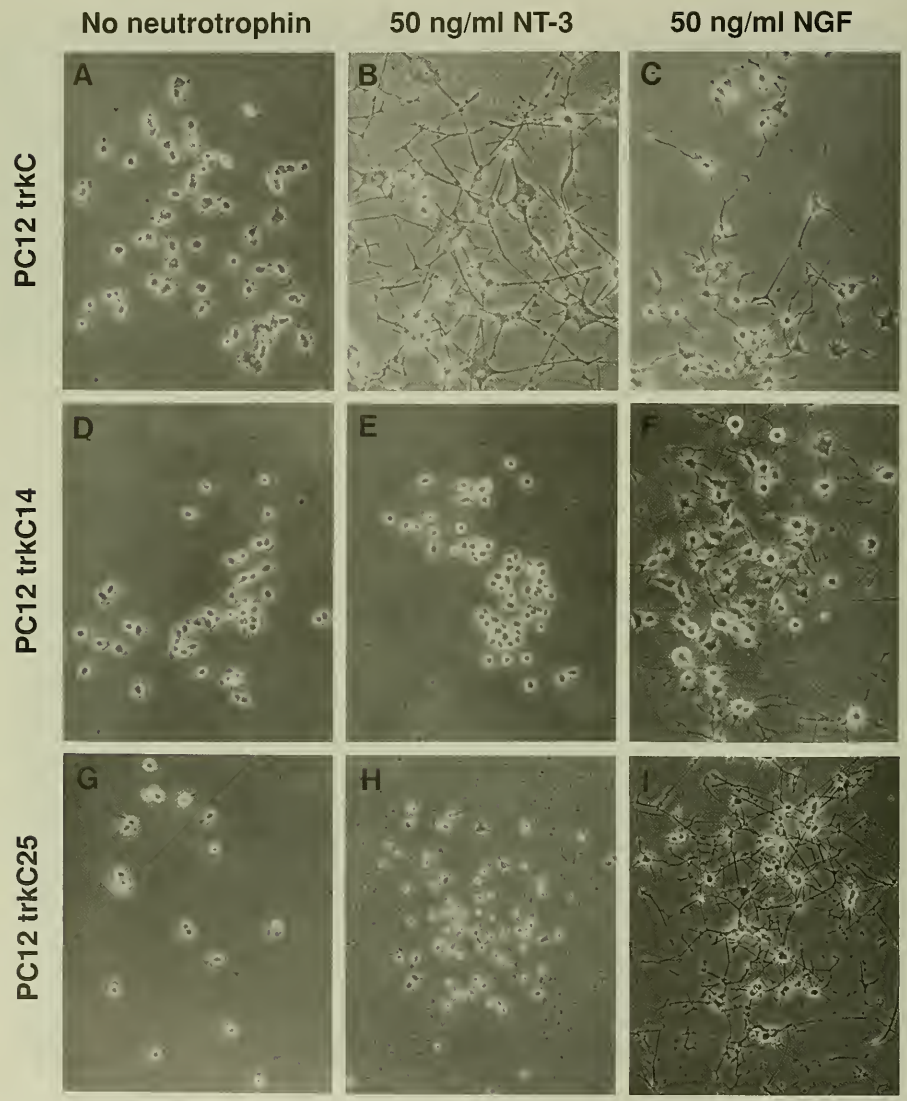


FIGURE 10. PC12 cell neurite outgrowth assay. PC12 cell lines expressing *trk*C (A, B, and C), *trk*C14 (D, E, and F), and *trk*C25 (G, H, and 1) were plated on collagen-coated dishes. The cells were treated with PC12 medium alone (A, D, and G), PC12 medium supplemented with 50 ng/ml NT-3 (B, E, and H), or PC12 medium supplemented with 50 ng/ml NGF (C, F, and I). Representative fields were photographed 3 days after treatment.

concentrations, with a half-maximal concentration of 50 pg/ml, the *trk*C14- and *trk*C25-expressing PC12 cells were completely unresponsive in this assay, even at NT-3 levels that were 1000-fold higher than the saturation concentration for the TrkC receptor (100 ng/ml; Figure 10A to 10I). Taken together, these data indicate that the insertions present in TrkC14 and TrkC25 render these receptors inactive in the PC12 neurite outgrowth assay, even though NT-3 can stimulate tyrosine phosphorylation in these cells.

Activation of Trk receptors in response to appropriate neurotrophins results in a series of recognizable downstream effects, including the association with and/or phosphorylation of substrates in the cytoplasm as well as the transcriptional activation of several previously identified genes. In collaboration with David Kaplan (Eukaryotic Signal Transduction Group, Molecular Mechanisms of Carcinogenesis Laboratory), we have assessed the state of several signal transduction intermediates after addition of NT-3 to *trk*C14- and *trk*C25-expressing PC12 cells. The data that are summarized in Table 2 indicate that the signaling pathways investigated are inactive for these two isoforms, while remaining active for TrkC and for the Trk receptor via NGF. Similarly, no evidence has been obtained that immediate early or late genes become activated in these cells in response to NT-3 addition (Table 3).

TABLE 2. Signal transduction substrates in the TrkC receptor pathway

	Rat TrkC		Rat TrkC14		Rat TrkC25	
	NT-3	NGF	NT-3	NGF	NT-3	NGF
PLC-γ1	+	+	–	+	–	+
ERKs	+	+	–	+	–	+
SNT	+	+	–	+	–	+

TABLE 3. NT-3 induction of genes in PC12 cells containing TrkC and related isoforms

	Rat TrkC		Rat TrkC14		Rat TrkC25	
	NT-3	NGF	NT-3	NGF	NT-3	NGF
Fos	+	+	–	+	–	+
Jun	+	+	–	+	–	+
NGF1-A	+	+	–	+	–	+
NGF1-B	+	+	–	+	–	+
Fra-1	+	+	–	+	ND*	ND
Fra-2	+	+	–	+	ND	ND

*Not determined.

Structure/Function Analysis

The availability of PC12 cell lines that express the various Trk receptors, and thus respond biologically to the different neurotrophins, provides a means for studying the effects of specific mutations in the neurotrophins. The availability of a high-resolution crystallographic map for NGF has allowed the generation of reliable computer models for the structures of the remaining neurotrophins. These models provide a framework for the prediction of which residues may be important for interaction with receptors. Using *trk*C-expressing PC12 cells, in collaboration with Leonard Pate and Roman Urfer (Genentech, South San Francisco, CA), we are investigating a series of NT-3 mutants designed to test these ideas. The biological activity of the various mutants (Figure 11) can be correlated with receptor binding capacity. These studies should provide useful information regarding the structurally important amino acids for NT-3 function as a ligand for TrkC. It is conceivable that this approach may also lead to the identification of neurotrophins with increased biological activity on Trk receptors.

To gain a better understanding of the residues in Trk family receptors that are important for ligand interaction, we have expressed cDNAs for the various receptors in *Xenopus* oocytes. In collaboration with Eugenio Santos (National Cancer Institute, Bethesda, MD), we previously showed that expression of Trk receptors in *Xenopus* oocytes, followed by NGF treatment, resulted in the induction of oocyte maturation as manifested by germinal vesicle breakdown (GVBD). More recently, we have expressed TrkB and TrkC receptors in oocytes and have obtained responses to neurotrophins that mirror the data previously obtained using N1H/3T3 fibroblasts (Table 4 and Figure 12). This oocyte injection system now provides an efficient means for rapidly screening in vitro-generated receptor mutants. Chimeric receptors between the different Trk family receptors, as well as point mutants, will be tested for activation and for induction of GVBD as a means of mapping the residues important for the interaction of the various neurotrophins with their respective receptors.

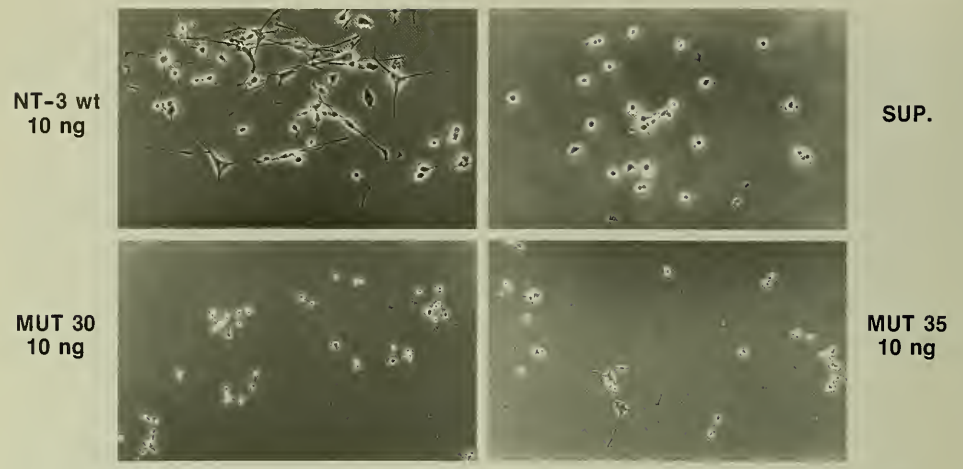


FIGURE 11. Analysis of mutated NT-3 molecules and their effects on *trk*C-expressing PC12 cells. Upper left: *trk*C-expressing PC12 cells treated with 10 ng wild-type NT-3. Upper right: *trk*C-expressing PC12 cells treated with control L93 cell supernatant. Bottom left: *trk*C-expressing PC12 cells treated with NT-3 mutant A (no differentiation). Bottom right: *trk*C-expressing PC12 cells treated with NT-3 mutant B (partial differentiation).

TABLE 4. Induction of meiotic maturation in injected oocytes

mRNA	None	NGF	BDNF	NT-3
None	–	–	–	–
trk	–	++	–	–
*trk*B	–	–	++	+
*trk*C	–	–	+	+

Oocytes were injected with *trk* family mRNA and maintained in 50% L-15 medium. Twenty-four hours after injection, oocytes were treated with neurotrophins (50 ng/ml) as noted for 15 hours and GVBD was scored (++ indicates > 90% GVBD; + indicates > 75% GVBD).

The Neurotrophins NT-3 and BDNF Enhance Neurogenesis in Cultures of Avian Neural Crest Cells

As previously noted, *trk*C is expressed in the very early stages of neural differentiation. In collaboration with James Weston (University of Oregon), we have tested the capacity of NT-3 and BDNF to support the survival of neural crest cells. Experiments performed on avian neural crest cells (Figure 13) indicate that exposure to BDNF or NT-3 results in increased numbers of neurons in subsequent cultures. These preliminary data do not permit discrimination between neurotrophin-mediated promotion of survival, differentiation, and/or proliferation of neural crest-derived neuronal precursors. Further experiments will be designed to distinguish between these alternatives, as well as to extend these studies to mouse neural crest.

FIGURE 12. *Xenopus* oocytes were injected with *trk*B mRNA and maintained in 50% L-15 medium. Twenty-four hours after injection, the oocyte on the right was treated with BDNF (50 ng/ml) for 15 hours. The oocyte on the left was untreated. The arrow indicates a white spot, which is indicative of GVBD.

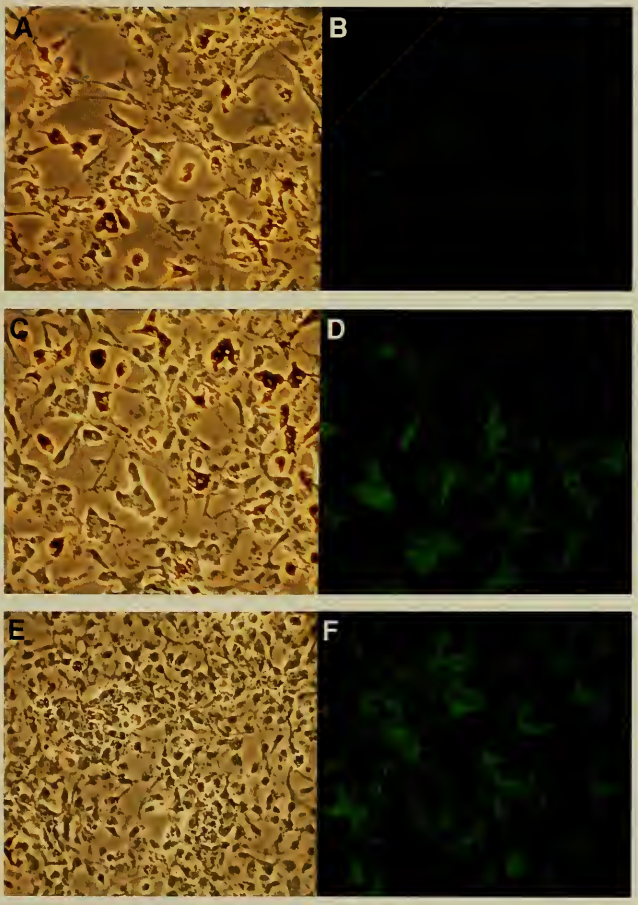


FIGURE 13. Phase-contrast (A, C, E) and fluorescence (B, D, F) photomicrographs of primary quail embryo neural crest cells after 5 days in vitro. When neural crest cell dispersal from the cultured neural tube is delayed by 32 hours in vitro, very few neurons, identified by the early neural marker, monoclonal antibody A2B5, differentiate (A, B). If the neural crest cells are exposed to NT-3 (C, D) or BDNF (E, F) during this delay in dispersal, the number of neurons in the cultures is increased.

Cell Biology of Development and Differentiation Group

Peter Donovan

Linzhao Cheng
Debra Compton
Susanna Dolci
Ganesan Gopalan
James Resnick
Dineli Wickramasinghe

Primordial germ cells (PGCs) are the embryonic precursors of the sperm and ova of the mature adult animal. As such, they are of great importance to the survival of animal species, since they are the cells responsible for carrying the genome from generation to generation. Notwithstanding this, until recently we have known very little about the factors regulating the development of the germ line, especially during embryogenesis. The elucidation of these factors is the primary goal of the Cell Biology of Development and Differentiation Group.

In the mouse, PGCs arise as a small population of cells in the postimplantation embryo at about 7 days post coitus (dpc). Arising initially within the ectoderm of the primitive streak, they move (or are moved) through the primitive streak until they reach the extraembryonic mesoderm at the caudal end of the embryo near the allantoic rudiment. At this time (approximately 8.5 dpc), PGCs are a population of about 100 cells that are classically identified by alkaline phosphatase (APase) histochemistry. Over the next day of development, they migrate to the embryo proper carried by the morphological movements of the embryo and come to lie within the hindgut endoderm. Here they begin the next phase of their development, active migration towards the gonad anlagen. Within 2 to 3 days, all the PGCs have colonized the gonad and their numbers have increased from the 100 cells that began migration to the 5,000 to 10,000 cells that populate the gonad by 12.5 dpc. PGCs therefore proliferate during migration and continue to do so for a few days after colonization of the gonad anlagen. By 13.5 dpc, approximately 25,000 PGCs can be identified by APase activity within the gonad; these cells eventually will form the meiotic population. In female embryos, PGCs enter directly into meiosis, arrest at meiotic prophase, and reenter meiosis directly after birth. In male embryos, on the other hand, PGCs enter a mitotic arrest (G_0 arrest) and will resume mitosis after birth. Male PGCs will enter meiosis at puberty, approximately 15 days after birth.

This brief sketch describes the embryonic history of the germ line in the mouse but contains little information about the molecules that regulate development. We have adopted two main approaches to this problem: firstly, to analyze peptide growth factors that may regulate PGC development and, secondly, to analyze cell cycle regulators that may control mitotic/meiotic transition or meiotic progression. The present report summarizes our data on the first approach.

Growth Factors That Affect Germ Cell Development

A major advance in our understanding of germ cell development came with the characterization of two mouse mutations that, when in the homozygous state, cause sterility. The mutations, Dominant White Spotting (*W*) and Steel (*Sl*), were shown to encode respectively a tyrosine kinase receptor (c-Kit) and its cognate ligand, variously termed c-Kit–ligand, stem cell factor, mast cell growth factor, and Steel factor (SLF). Previously, we reported on our studies on the role of SLF in PGC development. In the course of these studies, we have also identified other factors that regulate PGC proliferation.

At least two other growth factors have been identified that affect PGC development in the mouse embryo. These factors are leukemia inhibitory factor (LIF) and basic fibroblast growth factor (bFGF). While SLF is required for PGC survival in culture, it is not a powerful PGC mitogen. However, together with LIF, SLF can promote PGC proliferation. These two factors can therefore act as comitogens. Similarly, SLF and bFGF together stimulate increased PGC proliferation and could be considered comitogens. In the presence of binary combinations of factors, PGC proliferation in vitro is still finite and may be related to the proliferative period in vivo. Indeed, we have previously argued that PGCs develop on schedule in culture. However, in the presence of bFGF, SLF and LIF promote long-term proliferation of mouse PGCs and the derivation of established cell lines. These cell lines, which we have termed embryonic germ or EG cells, resemble embryonic stem (ES) cells, pluripotent cells derived from the preimplantation embryo, and embryonal carcinoma (EC) cells, pluripotent cells from germ cell-derived tumors. Given the similarity of EG cells to pluripotent EC and ES cells, we and others have tested their developmental potential in vitro and in vivo. These assays include the ability to form embryoid bodies in feeder-independent culture, to form teratomas in suitable histocompatible hosts, and to form chimeras when introduced into host blastocysts. In these assays, EC cells are developmentally pluripotent and give rise to tissues derived from all the three primary embryonic germ layers. ES cells, in addition, can contribute with high efficiency to the germ line. When tested in these assays by our group and by others, EG cells fall somewhat between ES and EC cells in that they are developmentally pluripotent and contribute to the germ line but at very low frequency. Creation of new EG cell lines from different strains of mice and further testing of their developmental potential may resolve this issue.

Germ Cells Express a LIF Receptor

This year, we have directed our efforts at understanding how these three factors — SLF, LIF, and bFGF — act to affect PGC survival, proliferation, and differentiation. A key question is whether these factors act directly on PGCs or whether they act indirectly via stromal cells or feeder cells. Genetic, biochemical, and immunological data clearly demonstrate that SLF acts directly on PGCs via the c-Kit tyrosine kinase receptor.

Previously, we showed that PGCs at 15.5 dpc express a low-affinity LIF receptor, termed LIFRβ (1992 Annual Report). This receptor, together with gp130 [originally identified as the interleukin-6 (IL-6) signal transducer], forms a high-affinity LIF receptor that is able to bind LIF and transduce signals in response to LIF. Using a polyclonal rabbit antibody to LIFRβ, we have shown that mitotic PGCs express LIFRβ. Genital ridges from 11.5-dpc embryos were disassociated and labeled with the TG-1 monoclonal antibody, which recognizes the stage-specific embryonic antigen-1 (SSEA-1). This carbohydrate differentiation antigen is a specific marker for PGCs in the mouse embryo from 8.5 to 13.5 dpc. By fluorescence-activated cell sorter (FACS) analysis, approximately 70% of SSEA-1$^+$ cells were also LIFRβ$^+$, thus confirming our earlier in situ hybridization data (Figure 1). We have also used the anti-LIFRβ

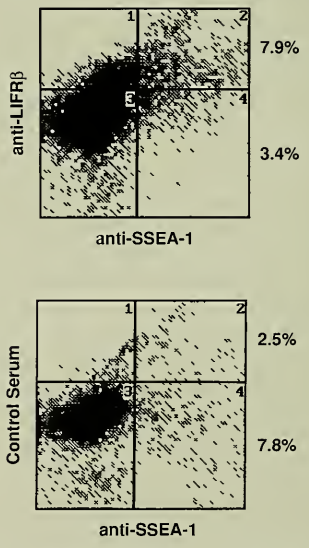


FIGURE 1. PGCs express a LIF receptor. The figure shows FACS analysis of PGCs isolated from 11.5-dpc genital ridges and stained with a germ cell-specific antibody (anti-SSEA-1). These populations were further stained with either an anti-LIFRβ serum or a control serum. The windows were arbitrarily set up to separate SSEA-1 and LIFRβ positive cells from the background.

antibody to block PGC proliferation in vitro. On feeder layers of STO cells (which produce LIF), LIFRβ antibodies completely inhibited PGC proliferation (Figure 2). This effect is comparable to data showing that antibodies to c-Kit inhibit PGC proliferation. Thus, signaling through the LIFRβ pathway is as important for PGC survival and proliferation as signaling via the c-Kit pathway. Antibodies to the IL-6 receptor (IL-6Rα) had no effect on PGC proliferation. Moreover, IL-6 is not mitogenic for PGCs, suggesting that signaling via the IL-6Rα/gp130 heterodimer receptor complex is unimportant for PGC proliferation. Although PGCs express LIFRβ, and signaling via the LIFRβ/gp130 heterodimer complex is required for PGC proliferation, we and others have been unable to detect LIF expression in the developing gonad. Moreover, mice lacking LIF are fertile and have normal numbers of PGCs. These data suggest either that LIF is not the physiological ligand for PGCs or that in the absence of LIF another factor can duplicate its function. LIF is a member of a cytokine family that includes IL-6, interleukin-11 (IL-11), ciliary neurotrophic factor (CNTF), and oncostatin-M (OSM). These factors are related by secondary structure, but not by primary amino acid structure. In addition to structural similarity, two of the factors, OSM and CNTF, can bind to the LIFRβ/gp130 heterodimer complex. While OSM can bind directly to the heterodimer complex, CNTF requires the presence of a third receptor component, CNTFRα, and thus has a tripartite receptor.

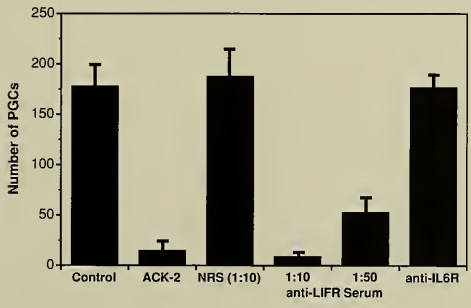


FIGURE 2. Anti-LIFRβ serum blocks PGC proliferation. PGCs were seeded onto STO cell monolayers and exposed to anti-c-Kit (ACK-2) serum, anti-LIFRβ serum, or a monoclonal anti-IL-6R antibody. Antibodies to c-Kit and LIFRβ block PGC proliferation.

Using purified rat CNTF and human OSM, we have shown that these factors can stimulate PGC proliferation (Figure 3). We have established a hierarchy of mitogenic activity of these cytokines for PGCs. LIF and OSM are equivalent in their activity: CNTF is less active, followed by IL-11 (not shown). In our hands, IL-6 had no direct effect on PGC proliferation, thereby confirming data presented above that signaling via the IL-6Rα/gp130 heterodimer complex is not required for PGC proliferation.

These data suggest that in LIF-deficient animals, other members of the LIF/IL-6 cytokine family duplicate the function of LIF. An alternative hypothesis is that cytokines such as OSM or CNTF may be the natural physiological ligand for PGCs. Key questions are whether CNTF or OSM are expressed in the developing gonad and whether, in the case of CNTF, its receptor (CNTFRα) is expressed on the PGC surface. Molecular probes for CNTF and CNTFRα are available and should allow this question to be answered shortly. However, OSM has only been identified in humans and simians, and a mouse homolog has not yet been identified. However, taken together these data suggest that LIF or a related cytokine can act directly on PGCs to stimulate PGC proliferation.

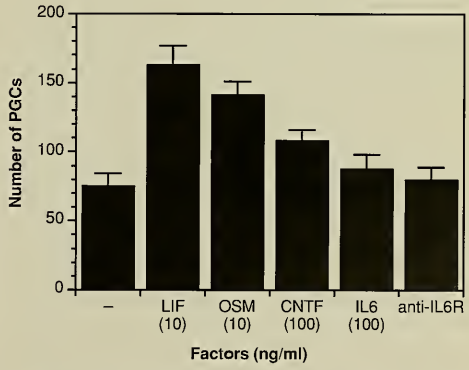


FIGURE 3. Effect of LIF and some related cytokines on PGC proliferation in culture. PGCs were seeded on gelatin-coated wells in the absence of STO feeder layers and incubated with LIF, OSM, CNTF, or IL-6 at the given concentrations or with monoclonal antibody against IL-6R (50 μg/ml). Each bar represents the mean and standard deviation of three wells.

Mode of Action of bFGF

The mechanism of action of bFGF has proven more difficult to define in part because of the complexity of FGF receptors and because of the large number of types of FGFs. At present, at least nine types of FGF have been identified: acidic FGF (FGF-1), basic FGF (FGF-2), int-2 (FGF-3), Kaposi FGF/hst (FGF-4), FGF-5, FGF-6, keratinocyte growth factor (FGF-7), FGF-8, and FGF-9. Each of these has a distinct expression pattern and binding properties. In addition, four major FGF receptors have been identified, which are members of the receptor tyrosine kinase superfamily. Like the factors, these receptors, termed FGFR-1 (*flg*), FGFR-2 (*bek*), FGFR-3, and FGFR-4, each has a distinct expression pattern and affinity for different FGFs. Previous studies by our group and others have clearly shown that bFGF can stimulate fibroblast feeder cells to produce LIF in two forms, a matrix-associated form (LIF-M) and a diffusible form (LIF-D). Therefore, bFGF could stimulate PGC proliferation by stimulating the feeder layers to produce LIF. Given that exogenous, soluble LIF (equivalent to LIF-D) is probably present in saturating amounts in these feeder layers, it seems unlikely that bFGF acts via its ability to upregulate LIF-D. It is conceivable, however, that upregulation of LIF-M could significantly alter PGC proliferation rates. However, other forms of FGF, such as FGF-4, can stimulate LIF production by fibroblast feeder cells without concomitant increases in PGC proliferation. These data suggest that bFGF also acts directly by interacting with an FGF receptor expressed by PGCs.

Moreover, bFGF can stimulate PGC proliferation on other fibroblast feeder layers without concomitant stimulation of LIF production. For example, bFGF can stimulate PGC proliferation on bone marrow-derived stromal cells as well as on STO cells (Figure 4). However, LIF levels are not altered by bFGF stimulation, although levels of SLF are increased. In certain conditions, PGCs can be maintained in the absence of feeder cells for short periods of time. When 11.5-dpc PGCs are isolated from genital ridges and plated onto gelatin-coated dishes, they will survive for up to 3 days and proliferate in response to exogenous growth factors and serum. When these cells are exposed to bFGF, they proliferate for 2 days. Since the number of contaminating somatic cells is relatively small, it seems unlikely that in these conditions bFGF is acting indirectly.

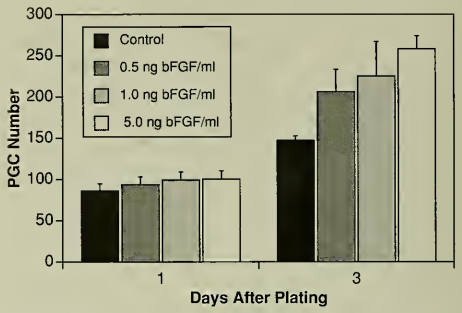


FIGURE 4. PGC proliferation on bone marrow-derived cell lines. PGCs obtained from the caudal portion of 8.5-dpc embryos were plated on irradiated BM-WT feeder cells in control medium (DMEM containing 15% FCS) or in the same medium containing the indicated concentrations of human recombinant bFGF. PGC numbers were determined following alkaline phosphatase staining. Each bar represents the mean and standard deviation of four wells.

Taken together, these data suggest that FGF may act directly on PGCs rather than indirectly via the fibroblast feeder layer. A crucial question is whether PGCs express an FGF receptor. It is becoming apparent that each FGF receptor will exhibit a distinct hierarchy of binding affinities towards FGF ligands. For example, a form of the FGFR-1 (*flg*) binds FGF-1 and FGF-2 with 15-fold higher affinity than FGF-4.

In order to understand a potential role for bFGF in PGC proliferation, we have sought to understand the expression pattern of bFGF (and other FGF family members) as well as the FGF receptors. In situ hybridization analysis has previously been carried out to determine the expression pattern of some FGFs during development. FGF-4 has been detected in the egg cylinder at the time at which PGCs are first identifiable, and both FGF3 and FGF5 are expressed in the primitive streak and could potentially act on PGCs. Data are available on the expression of FGFs on the PGC migratory route or in the developing gonad at later stages of development. We have used sensitive RNase protection assays to identify FGF receptor transcripts in the developing gonad. So far, using this technique we have identified both FGFR-1 (*flg*) and FGFR-2 (*bek*) in the gonad, although we cannot determine whether their presence is due to expression by PGCs (Figure 5). Presently, we are carrying out in situ hybridization analysis to answer this question.

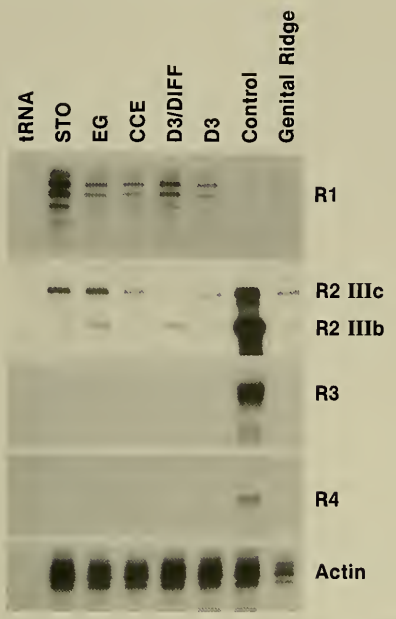


FIGURE 5. RNase protection analysis of FGF receptor expression in cell lines and tissues. RNase protection was carried out on RNA isolated from STO, EG, CCE (ES), and D3 (ES) or their differentiated derivatives as well as genital ridge and skin (control) RNA. An actin probe was used as a control for loading. Four receptors were analyzed: R1 (*flg*), two splice variants of R2 (*bek*) (IIIb and IIIc), R3, and R4.

Modification of Germ Cell Differentiation

Since all three factors (SLF, LIF, and bFGF) act to alter the developmental pathway of PGCs, it is important to understand how these factors act in vivo. Presumably, if PGCs were to be exposed to all three factors in vivo, it would alter their differentiation, leading to tumorigenesis and possibly sterility. Two possibilities seem likely. First, PGCs never encounter all three factors in vivo and one factor (probably LIF or bFGF) can alter their developmental pathway. In this scenario, LIF or bFGF would be produced aberrantly or as a result of tissue damage. Second, PGCs do encounter all three factors but their combined effect is attenuated partly by another negative regulator. Such a factor could be transforming growth factor (TGFβ1), which has been shown to be present in the developing gonad, or Mullerian inhibitory substance (MIS), another member of the TGFβ superfamily known to be involved in gonadal development.

Once PGCs have been exposed to SLF, LIF, and bFGF and have formed EG cells, they will grow indefinitely in culture and therefore could be considered immortalized. Moreover, we and others have shown that EG cells become growth factor independent once established. Therefore, EG cells can be placed on feeder layers of STO cells or primary mouse embryo fibroblasts without added SLF, LIF, or bFGF. These data suggest that part of the mechanism of EG cell formation involves establishment of autocrine loops for some, if not all, of these factors. To date, we have found that an FGF autocrine loop may be established in EG cells. Conditioned medium from EG cells grown on STO feeder cells will stimulate the proliferation of mouse NIH/3T3 cells, while conditioned medium from the STO feeder cells alone has no effect. This result suggests that EG cells produce a form of FGF.

In order to study FGF expression by EG cells, we have developed a method for isolating EG cells from STO feeder layers (or other fibroblast feeder layers) by unit gravity sedimentation. This procedure involves trypsinization of cultures at room temperature. Under these conditions, the trypsin dissociates the single-cell feeder layer but is unable to dissociate the large groups or clumps of EG cells. This mixture of single cells and clumps is then overlaid on a column of medium with serum. In these circumstances, the larger, denser clumps of EG cells settle more rapidly than the single cells and therefore can be harvested from the bottom of the column after 5 to 10 minutes. With two or three rounds of unit gravity sedimentation, it is possible to obtain highly purified populations of EG cells for RNA isolation.

Using RNA isolated from EG cells purified in this way, we have begun to analyze FGF and FGF receptor expression in these cells. So far, we have identified FGF-4 (Kaposi FGF/hst) RNA and RNA for a number of FGF receptors in EG cells. FGF-4 expression is of interest because this form of FGF is expressed by many EC cell lines (including F9 EC cells) and by many ES cells (including the D3 and CCE lines). Thus, FGF-4 is a candidate for an autocrine stimulator of EG cell growth.

In order to test the idea that FGFs could act as autocrine stimulators of EG cell growth, we have sought to interrupt this signaling pathway. Since effective neutralizing antibodies to FGF-4 are not yet readily available, we have used the drug suramin to interfere with FGF signaling. Suramin is a polyanion that acts to prevent productive interaction of heparan-binding growth factors with their receptors. We have observed that suramin reduces the growth rate of preexisting EG cell colonies. That this effect is due to interference with an FGF signaling pathway is evidenced by the partial ability of bFGF to overcome this effect. Although exogenous FGF can only partially overcome suramin inhibition, it is possible that EG cells produce more than one type of FGF. RNase protection analysis demonstrates that EG cells express two splice variants of FGFR-2 and may also express FGFR-1. Additionally, the inability of bFGF to completely overcome suramin inhibition of EG cell growth may reflect other effects of suramin. These effects could include decreased adhesion of EG cells to the feeder layers, toxicity of suramin, or interaction with non-FGF growth factors.

We have also investigated possible autocrine stimulation of EG cells by SLF acting through the c-Kit tyrosine kinase receptor. Many EC and ES cell lines express c-Kit and some ES cell lines express SLF. These results suggest that autocrine stimulation via the SLF/c-Kit axis is important for the maintenance of pluripotent stem cells. To investigate this possibility, we have used a variety of mouse fibroblast feeder cells, including cell lines derived from mice carrying mutations at the *Sl* locus. In addition, we have used CV-1 cells (derived from African green monkey kidney cells), which do not express SLF, and recombinant derivatives expressing SLF. These cells lines were produced by transfecting CV-1 cells with constructs capable of expressing a full-length form of SLF. On cell lines expressing full-length transmembrane forms of SLF, EG cells grow efficiently and form colonies. On cell lines not expressing SLF, EG cells do grow, but their growth is slow and colony size and number are reduced.

The factors regulating the growth of PGCs and EG cells will continue to be of interest, particularly since the process of EG cell formation is so similar to the formation of germ cell-derived tumors (teratomas). Using mouse mutants deficient in c-Kit and SLF, we will attempt to derive EG cell lines that should allow an analysis of the role of these molecules in EG cell derivation. Similarly, the availability of LIF-deficient animals should allow the role of LIF in EG cell derivation to be assessed. As mice lacking other factors or receptors become available, these animals will become invaluable tools in this research. The availability of pluripotent stem cell lines derived from PGCs has important consequences for the analysis of PGC development. In particular, these cell lines should allow analysis of germline imprinting and pluripotency. Moreover, the derivation of stem cells directly from PGCs represents a novel route that may be applicable to other experimental organisms and agriculturally important animals.

Molecular Mechanisms of Carcinogenesis Laboratory

George F. Vande Woude

Stephen H. Hughes
George N. Pavlakis
Barbara K. Felber
Peter F. Johnson
Deborah K. Morrison
David R. Kaplan

The seven principal investigators of the Molecular Mechanisms of Carcinogenesis Laboratory use molecular approaches to study gene regulation, retroviral replication, and the molecular basis of neoplastic transformation. The laboratory consists of three research sections and four research groups.

The Molecular Oncology Section, headed by George F. Vande Woude, investigates the molecular basis of neoplastic transformation by the *mos* and *met* oncogenes. They have found that the *mos* proto-oncogene is a regulator of oocyte maturation: Mos is necessary and sufficient for initiation of meiosis I and, as part of cytostatic factor (CSF), for the arrest of oocytes at metaphase II. Their examination of somatic cells that overexpress or are transformed by *mos* have shown that these cells express molecular and biochemical markers usually restricted to the meiotic cell cycle. These observations have led to the proposal that the characteristic phenotypes of transformed cells may result from the superimposition of the meiotic program onto the normal somatic cell cycle program. They found that the *met* proto-oncogene encodes a receptor tyrosine kinase, the ligand of which is hepatocyte growth factor/scatter factor (HGF/SF). Met is expressed at high levels in normal cells bordering lumen and ductal structures in vivo; while Met expression is greatly reduced in affected tumor cells of breast tumor biopsies. They also have shown that coexpression of Met and HGF/SF in NIH/3T3 cells renders these cells highly tumorigenic in nude mice by an autocrine mechanism and their recent studies have implicated Met-HGF/SF coexpression in the development of human soft tissue sarcomas. They have initiated studies to test whether oncogenes and tumor suppressor genes render cells vulnerable to antineoplastic drugs.

The Gene Expression in Eukaryotes Section, directed by Stephen Hughes, has focused on the structure and function of cellular and viral genes. In collaboration with Edward Arnold (Rutgers University), the high-resolution structure of human immunodeficiency virus type 1 (HIV-1) reverse transcriptase (RT) has been determined. They are complementing these structural studies with the genetic and biochemical analysis of over 250 specific RT mutations, including drug-resistant variants. They also are conducting biological and biochemical analyses of the *ski* oncogene, which

they found can cause a profound increase in the skeletal musculature of transgenic mice. They have developed a series of replication-competent avian retroviruses that can be used to express genes both in vitro and in vivo. Recently, these in vivo analyses have been extended to mammalian systems by their creation of transgenic mouse lines that express the cellular receptor for these retroviral vectors.

The Human Retrovirus Section, headed by George Pavlakis, studies the molecular biology and pathogenic mechanisms of the human retroviruses HIV and human T-cell leukemia virus (HTLV). They have focused on determining the mechanism of action of the HTLV Tax and the HIV/HTLV Rev/Rex regulatory proteins as well as elucidating the cis-acting elements that mediate Rev function and negatively modulate expression of the viral structural proteins. In collaboration with the Human Retrovirus Pathogenesis Group, they have characterized a family of interferon-inducible genes that encode RNA-binding proteins, some of which inhibit HIV-1 expression. These studies have clarified virus-cellular protein interactions and defined a novel mechanism for the antiviral activity of interferon, providing a basis for new antiviral strategies to combat HIV infection.

The Human Retrovirus Pathogenesis Group, led by Barbara Felber, is interested in the mechanisms of HTLV-induced carcinogenesis. They have identified viral-encoded accessory proteins (Rof and Tof) and are evaluating the role of the structural, regulatory, and accessory proteins in HTLV infection and neoplasia. They also are examining the posttranscriptional regulation of HTLV and HIV gene expression and the interplay between viral and cellular factors involved in these regulatory steps. These studies serve to elucidate viral functions that appear to rescue both viral and cellular defects in posttranscriptional RNA processing. They also are designing novel retroviral vectors with potential utility in gene therapy for HIV infection and cancer.

The Eukaryotic Transcriptional Regulation Group, headed by Peter F. Johnson, studies a subfamily of basic:leucine zipper proteins, the C/EBP transcriptional activator family. These proteins are DNA binding, show dimerization specificities, and are expressed in tissue-restricted patterns. Their group examines the role of these proteins in regulating cell-specific transcription during differentiation. For example, the expression of C/EBP-related protein-2 (CRP2) is coupled to cell cycle arrest and terminal differentiation of macrophages. They also have identified a negative regulatory domain of CRP2 and are elucidating the transcriptional and translational control of *c/ebp* and *crp2* gene expression.

The Cellular Growth Mechanisms Group, under the direction of Deborah Morrison, studies the proteins involved in regulating the cell growth process, specifically the Raf family of serine/threonine protein kinases. The *raf* genes were originally identified as oncogenes, and activated Raf proteins have been found in a variety of human and rodent tumors. Their group has demonstrated that Raf-1 plays an important role in transmitting proliferative, developmental, and oncogenic signals and are elucidating the mechanism by which oncogenes and mitogens activate Raf-1. Their biochemical studies of Raf-1 have shown that phosphorylation both positively and negatively modulates Raf-1 function.

The Eukaryotic Signal Transduction, headed by David Kaplan, investigates the mechanism of signal transduction during mitogenesis and cell differentiation. They have focused on signal transduction by nerve growth factor (NGF), which is required for the survival and differentiation of specific populations of neurons, and have identified the *trk* proto-oncogene as the receptor for NGF. In addition, they have identified several components of the signal transduction pathways used by NGF and the NGF family of neurotrophins. These studies led to the discovery of SNT, the first specific target of a differentiation factor-induced tyrosine kinase activity. They also have shown

that the morphogen retinoic acid (RA) is a specific inducer of the receptor encoded by the *trk*-related gene, *trk*B. Thus, RA induces *trk*B in the presence of its cognate ligand that leads to the differentiation of neuroblastoma cells. This observation may prove useful in the treatment of neuroblastoma tumors.

Molecular Oncology Section

George F. Vande Woude

Kristina Borror
Taesaeng Choi
Janelle Cortner
Donna Faletto
Cara Fiore
Kenji Fukasawa
Michael Jeffers
Han-mo Koo
Wayne Matten
Mary Jane McWilliams
Barbel Meissner
Linda Miller
Monica Murakami
Marianne Oskarsson
Sing Rong
Shen Rulong
Nicholas Schulz
Ilan Tsarfaty
Nelson Yew
Renping Zhou

For the past several years, we have been determining how the c-*mos* proto-oncogene functions during oocyte maturation. During the past year, we have investigated the effects of *mos* expression in somatic cells. We suggested many years ago that the phenotype of *mos*-transformed cells could be derived from the inappropriate expression of M-phase phenotypes during interphase of the somatic cell cycle. Two studies performed this year support this hypothesis and implicate M-phase cytoskeletal modifications in the expression of the transformed phenotype. First, in a collaborative study with Jonathan Cooper's laboratory (Fred Hutchinson Cancer Research Center, Seattle, WA) and Natalie Ahn (University of Colorado, Boulder, CO), we have shown that Mos appears to activate the kinase that activates mitogen-activated protein kinase (MAPK) and, therefore, Mos may be considered a MAPK kinase kinase. MAPK activation appears to be a downstream signaling event that is affected by many oncogenes, and may be a key in triggering expression of the phenotypes of transformed cells that are induced by oncogenes. Second, as first reported 25 years ago, Mos overexpression in somatic cells infected with Moloney murine sarcoma virus (Mo-MSV) containing v-*mos* causes cells to round up, detach from the monolayer, arrest growth, and become binucleated. We have shown that these floating cells display properties that occur in oocytes during meiosis. For example, we have shown that MAPK is activated to very high levels both during meiosis and in the floating cells. Activated MAPK has been implicated in microtubule reorganization during oocyte maturation and has recently been shown to localize to microtubule organizing centers. We have previously proposed that Mos is also implicated in microtubule reorganization (1991 Annual Report). Thus, we believe that one explanation for many of the phenotypes of transformed cells is that *mos* and other oncogenes that activate MAPK induce a program of meiotic cytoskeletal reorganization that runs concurrently with all stages of the somatic cell cycle.

We continue to make progress in understanding *met* proto-oncogene function and its activity as an oncogene. We discovered several years ago, together with the laboratory of Stuart Aaronson (National Cancer Institute, Bethesda, MD), that hepatocyte growth factor/scatter factor (HGF/SF) is the ligand for the *met* proto-oncogene product. Normal Met expression is restricted predominantly to epithelial cells, while HGF/SF is expressed in cells of mesenchymal origin. A variety of responses — mitogenic, motogenic, and morphogenic — have been attributed to this ligand/receptor signaling system and these responses generally are paracrine mediated. We reported last year that Met is involved in epithelial lumen formation, a phenotype that is most likely related to the motogenic and morphogenic responses (1992 Annual Report). Last year, we also reported that coexpression of Met with its ligand HGF/SF causes NIH/3T3 cells to become highly tumorigenic in nude mice via an autocrine mechanism.

Quite surprisingly, histopathological examination of these tumors reveal, in addition to the fibrosarcomas expected to arise from NIH/3T3 fibroblast cells, carcinoma-like focal areas that contain intercellular adhering junctions that are desmosomal-like. Moreover, as expected, the cells express vimentin, but they also express epithelial cell-specific cytokeratins. We have also found that Met is expressed in the developing embryonic kidney, in cells that undergo mesenchymal to epithelial cell conversion to form the tubules (1992 Annual Report). These embryonic cells express both HGF/SF and Met, as well as vimentin and cytokeratin. These studies suggest that Met:HGF/SF contributes to mesenchymal cell differentiation that leads to the formation of epithelial cells. This discovery can help to explain the origin of a variety of tumors that express both mesenchymal and epithelial markers. In addition, adding to our discovery that Met expression in NIH/3T3 cells is highly tumorigenic due to an autocrine mechanism, we have discovered that Met and HGF/SF are coexpressed in human soft tissue sarcomas and may contribute to the formation of these tumors.

Mos and the Molecular Basis of the Transformed Phenotype

The product of the *mos* proto-oncogene is a serine/threonine kinase that is expressed at high levels in germ cells. In previous years, we have shown that Mos is a regulator of the program of meiotic maturation that results in the production of an unfertilized egg. Mos is also a component of cytostatic factor (CSF), an activity that is believed to arrest oocyte maturation in meiotic metaphase II. We have shown that Mos is associated with tubulin in unfertilized eggs and transformed cells (1990 Annual Report), raising the possibility that Mos is involved in the microtubular reorganization that occurs during M-phase. We have proposed that inappropriate expression of this M-phase meiotic activity during interphase of the cell cycle is responsible for the transforming activity of Mos.

A mitogen-activated protein kinase (MAPK), Xp42, is also activated during oocyte maturation in *Xenopus*. This kinase is present in stage VI oocytes and is activated about 2 hours after the onset of Mos protein synthesis and contemporaneously with the activation of maturation promoting factor (MPF; $p34^{cdc2}$/cyclin B). MAPK is dephosphorylated approximately 20 minutes after egg fertilization, about the time of Mos protein degradation, and MAPK is not rephosphorylated during early embryogenesis. MPF activity is reactivated in subsequent mitotic cell cycles, although neither Mos nor MAPK is reactivated during embryogenesis in *Xenopus*. Hence, the activities of Mos and MAPK appear to be coordinately regulated throughout oocyte maturation and early embryogenesis. We were interested in determining whether any functional interaction exists between Mos and MAPK.

Mos Activates MAPK in *Xenopus* Oocytes. These studies were performed in collaboration with Drs. Cooper and Ahn. Maltose binding protein (MBP)-Mos fusion proteins were injected into oocytes to determine whether Mos kinase activity had any effect on the *Xenopus* MAPK Xp42. The phosphorylation state of MAPK was monitored by its migration on SDS-polyacrylamide gels and its catalytic activity was examined by in vitro kinase assays (Figure 1). We found that injection of wild-type Mos fusion protein resulted in rapid activation of MAPK, whereas the kinase-inactive Mos did not activate MAPK. Mos-induced activation of MAPK resulted in phosphorylation of approximately 80% of the MAPK molecules by 1 hour after injection. In contrast, only 10% to 20% of the MAPK was phosphorylated after 1 hour when oocytes were treated with progesterone in the absence of cycloheximide (data not shown), but 100% phosphorylation occurred after 6 hours, at which time germinal vesicle breakdown (GVBD) was virtually complete. The catalytic activity of the phosphorylated MAPK in MBP-Mos–injected oocytes at both 1 hour and 6 hours was comparable to that observed in progesterone-matured oocytes after 6 hours. In addition, this phosphorylation appeared to be carried out via a *Xenopus* MAPK kinase (MAPKK), since extracts of oocytes injected with wild-type Mos activated MAPKK and were able to phosphorylate a recombinant ERK2 isoform of MAPK (Figure 2). In contrast, the extracts of oocytes injected with kinase-inactive MBP-Mos (KM) were not able to phosphorylate ERK2.

These results suggested that MPB-Mos could rapidly activate MAPK and a MAPKK, in the absence of protein synthesis. We further compared the kinetics of activation of MAPK and MPF. Oocytes were injected with MBP-Mos in the absence of both cycloheximide and progesterone.

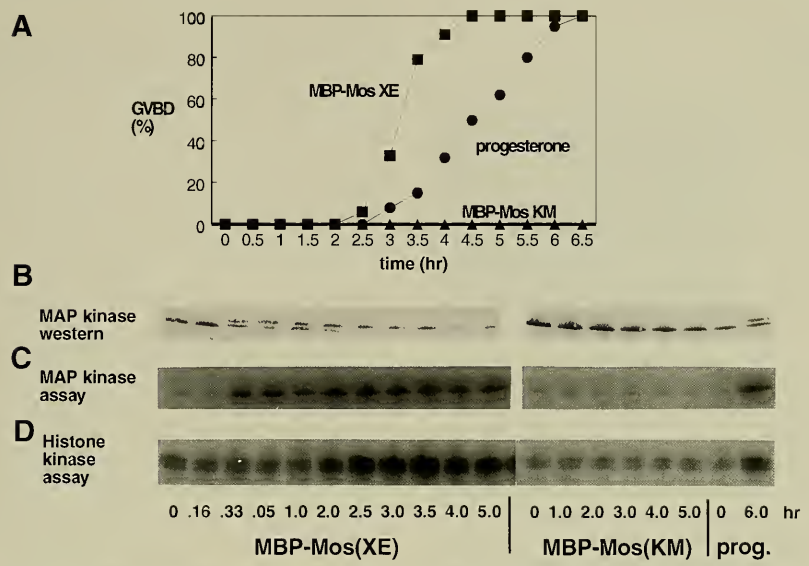


FIGURE 1. Kinetics of activation of MAPK and MPF. Oocytes were given injections of wild-type (XE) or kinase-inactive (KM) MBP-Mos fusion proteins or were incubated with progesterone. Groups of 10 oocytes were homogenized at each time point. MAPK was immunoprecipitated, and in vitro kinase assays were done on the immunoprecipitates with GST-Myc as a substrate. To measure MPF activity, $p34^{cdc2}$ was affinity purified with $p13^{suc1}$-agarose and in vitro kinase assays were performed with histone H1 as a substrate. (A) Percentage of oocytes showing GVBD. (B) Immunoblot of immunoprecipitated Xp42 with antibody to Xp42. (C) Phosphorylation of GST-Myc by immunoprecipitated Xp42. (D) Histone H1 phosphorylation by $p13^{suc1}$-associated kinases.

Under these conditions, MBP-Mos induces complete maturation (1992 Annual Report). Control oocytes were not injected, but were incubated with progesterone. Progesterone induced 100% GVBD within 6.5 hours and 50% GVBD by 4.5 hours (Figure 1A), while the wild-type MBP-Mos induced 100% GVBD within 4.5 hours and 50% GVBD by 3.2 hours. Injection of wild-type MBP-Mos induced phosphorylation of about 50% of the Xp42 (Figure 1B), and stimulated MAPK catalytic activity within 20 minutes following injection (Figure 1C). Phosphorylation of Xp42 was complete by 2 hours, and catalytic activity remained high throughout the time course. In contrast, MPF, measured by $p34^{cdc2}$-associated histone H1 kinase activity, was not elevated until approximately 2 hours following injection of the wild-type MBP-Mos (Figure 1D). Kinase-inactive MBP-Mos (KM) was not able to induce GVBD, MAPK activation, or MPF activity.

Mos Is Able to Activate a Purified MAPKK in Vitro. We tested the ability of Mos to phosphorylate and activate a MAPKK in vitro. Wild-type and kinase-inactive MBP-Mos were treated by incubation with rabbit reticulocyte lysate (to activate wild-type MPB-Mos) and purified by immunoprecipitation. Mos kinase activity was evidenced by autophosphorylation of the wild-type but not the kinase-inactive MBP-Mos (Figure 2, lanes 1 and 2). Purified, active MAPKK from rabbit skeletal muscle was able to phosphorylate ERK2 in the absence of MBP-Mos [(Figure 2, lane 10; note that the MAPKK is also a substrate for ERK2 (compare Figure 2, lanes 9 and 10)]. The phosphorylation of ERK2 by MAPKK was greatly decreased by prior incubation of MAPKK with protein phosphatase 2A, followed by treatment with the phosphatase inhibitor microcystin (lane 12). After incubation of phosphatase-treated MAPKK with wild-type, but not kinase-inactive, MBP-Mos, phosphorylation of

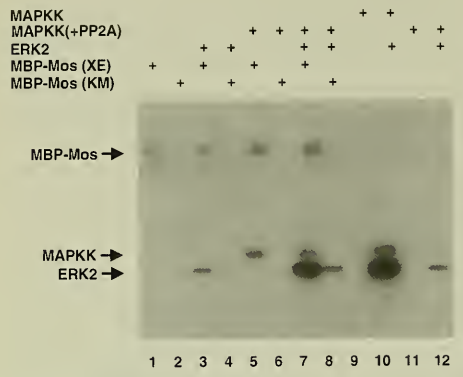


FIGURE 2. MBP-Mos phosphorylates and activities a MAPKK. Wild-type (XE) or kinase-inactive (KM) MBP-Mos fusion proteins were activated by incubation with rabbit reticulocyte lysate in the presence of an ATP-generating system and then repurified by immunoprecipitation. MAPKK purified from rabbit skeletal muscle was inactivated by treatment with protein phosphatase 2A (PP2A). The MBP-Mos immunoprecipitates were incubated in a kinase cocktail containing [γ-^{32}P]ATP (lanes 1 and 2); the arrow indicates autophosphorylation of the MBP-Mos (XE). The immunoprecipitates were mixed with recombinant ERK2 (lanes 3 and 4); the arrow indicates a small amount of phosphorylation of ERK2 by MBP-Mos (XE) and a lesser amount of autophosphorylation of ERK2 in the presence of MBP-Mos (KM). In lanes 5 and 6, the Mos immunoprecipitates were mixed with PP2A-inactivated MAPKK; the arrow indicates phosphorylation of the MAPKK. In lanes 7 and 8, the immunoprecipitates were mixed with inactivated MAPKK and ERK2; the arrow indicates phosphorylation of the MAPKK as well as ERK2. In lanes 9 to 12, the MBP-Mos immunoprecipitates were omitted, and untreated (lanes 9 and 10) or PP2A-treated (lanes 11 and 12) MAPKK was incubated in the absence (lanes 9 and 11) or presence (lanes 10 and 12) of ERK2.

MAPKK was detected (lanes 5 and 6). This phosphorylation was presumably on the regulatory residue(s), since it markedly activated the phosphorylation of ERK2 (lanes 7 and 8). In the absence of MAPKK, wild-type MBP-Mos phosphorylated ERK2 only weakly (lane 3), and only trace ERK2 autophosphorylation was evident with kinase-inactive MBP-Mos (lane 4). The MBP-Mos phosphorylation of MAPKK apparently proceeded to high stoichiometry, since this MAPKK phosphorylated the ERK2 substrate nearly as well as fully active MAPKK (compare lanes 7 and 10). We do not know the efficiency of the reaction, however, because the rates of phosphorylation of MAPKK by MBP-Mos were not measured.

Mos-Induced Mitotic Arrest Results in MAPK Phosphorylation. We showed that Mos is the active component of CSF, and injecting Mos into one blastomere of a two-cell embryo has the dramatic effect of arresting mitotic divisions in that blastomere (1990 Annual Report). We were interested in learning whether the signaling pathway activated by Mos and present in the unfertilized egg is intact following fertilization. Therefore, we injected MBP-Mos fusion proteins into one blastomere of a two-cell embryo (Figure 3). After 30 minutes, when arrest of the injected blastomeres was evident, we immunoprecipitated Xp42 from the embryos. Arresting mitosis with wild-type Mos resulted in Xp42 phosphorylation and activation of catalytic activity (Figure 3, lane 3), while the kinase-inactive Mos did not activate MAPK (lane 2).

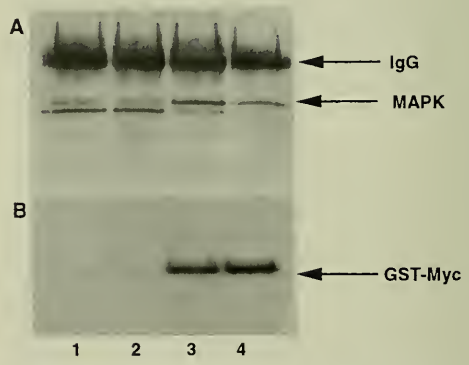


FIGURE 3. Mos-induced MAPK activation during mitotic arrest of embryos. Two-cell embryos were either untreated (lane 1) or given injections of kinase-inactive (lane 2) or wild-type (lane 3) MBP-Mos and incubated for 30 minutes. Extracts were prepared, and MAPK immunoprecipitated. As a positive control, a similar amount of egg extract was made and immunoprecipitated (lane 4). (A) Immunoblot of Xp42 immunoprecipitates probed with antibody to Xp42. IgG, immunoglobulin G. (B) In vitro kinase assays on the same immunoprecipitates with GST-Myc as a substrate.

These results show that injecting MBP-Mos into *Xenopus* activates MAPKK and MAPK prior to and apparently independently of MPF activation. These effects of Mos require a functional kinase domain and presumably occur as a result of the phosphorylation of one or more oocyte proteins after MBP-Mos injection. MBP-Mos in vitro reactivates MAPKK, suggesting that Mos may be a "MAP kinase kinase kinase," which may normally activate MAPK during oocyte maturation. Moreover, we found that MBP-Mos activates MAPK in *Xenopus* cleavage embryos concomitant with arrest at mitotic metaphase. Since Mos is normally present at meiosis but not mitosis, whereas MPF is present in both situations, Mos, not MPF, most likely governs MAPK activity at meiotic metaphase. MAPKKs have been shown to be activated by both normal and oncogenic mutant Ras, and MAPK activation by nerve growth factor, insulin, or platelet-derived growth factor requires the endogenous Ras protein. Likewise, Ras activates a MAPKK. Thus, MAPK activation appears to be a common signaling point at which Ras and other oncogenes (Raf and Src) converge, and this activation is possibly responsible for some of the common features of the cells they transform.

In vitro, MAPK phosphorylates a microtubule-associated protein and causes interphase microtubules to resemble metaphase microtubules. We have proposed that Mos is involved in the reorganization of microtubules that leads to spindle assembly (1990 Annual Report). It is possible that Mos, Ras, and other oncogenes, by activating MAPK, alter microtubules in transformed cells. The following studies show that the meiotic activities of Mos and MAPK may be responsible for many of the properties of transformed cells.

Somatic cells expressing Mos display properties that are similar to properties observed in maturing oocytes. As shown many years ago by Peter Fischinger (University of South Carolina Cancer Center, Charleston, SC), after acute infection of Swiss 3T3 cells with Mo-MSV, which contains the *v-mos* oncogene, cells expressing high levels of v-Mos round up and detach from the monolayer. Seventy percent of the floating cells exclude trypan blue, but do not proliferate. We have found that the floating cells are growth arrested with *2C* or *4C* DNA content (Figure 4) and both populations display inappropriate chromosome condensation (data not shown). The cells with *4C* DNA content are binucleated and have undergone nuclear division in the absence of cell division, a process reminiscent of the events of meiotic nuclear division (Figure 4A, panel c; Figure 4B, panels e and f).

We believe that the mononucleated cells are derived from cells that were in G_1 at the time of growth arrest, but the formation of binucleated cells requires a unique explanation. To account for binucleation in one-third of the floating cells, we assume that cells in G_2 through some stage of M-phase must have undergone karyokinesis, but not cytokinesis, before growth arrest occurred. Thus, Mos overexpression in Swiss 3T3 cells induced karyokinesis but not cell division. Similarly, in collaborative studies with Xiao Min Wang and Gary Borisy (University of Wisconsin, Madison, WI), we have found that the injection of *Xenopus* Mos into kangaroo rat epithelial cells at prometaphase induces karyokinesis in the absence of cytokinesis (data not shown). Somehow, Mos appears to interfere with somatic cell cytokinesis. We propose that this effect is due to expression of an activity associated with meiosis.

We have found other properties that are similar between the floating cells and maturing oocytes during meiosis, especially during meiotic interphase [the stage between meiosis I and II (Figure 5)]. Cyclin B is present and $p34^{cdc2}$ is dephosphorylated during this stage (data not shown), but histone H1 kinase activity is low and MAPK activity is very high (Figure 5A). Mos has been directly implicated in MAPK activation during meiosis (Figures 1 to 3), and the level of MAPK is significantly higher during meiosis and meiotic interphase than during mitosis in two-cell mouse embryos (Figure 5A). Also, during meiotic interphase, the chromosomes remain condensed even though MPF levels decrease (Figure 5B). The floating cells display these same properties — notably, MAPK is constitutively activated in Swiss 3T3 cells expressing v-Mos (Figure 6). MAPK activity was not detected in control Swiss 3T3 cells maintained for 36 hours in 0.5% serum (Figure 6, lane 1), but the floating cells and the v-*mos*–transformed cell lines (Tx1 to Tx3) displayed high levels of activity (Figure 6B, lanes 3 to 6). These results indicate that Mos expression in somatic cells, as in oocytes, is associated with the activation of MAPK. This activation is constitutive in 0.5% serum and therefore occurs in the absence of mitogenic stimulation.

Our results suggest a model for transformation by the *mos* oncogene whereby the program that is normally restricted to the highly ordered process of meiosis runs

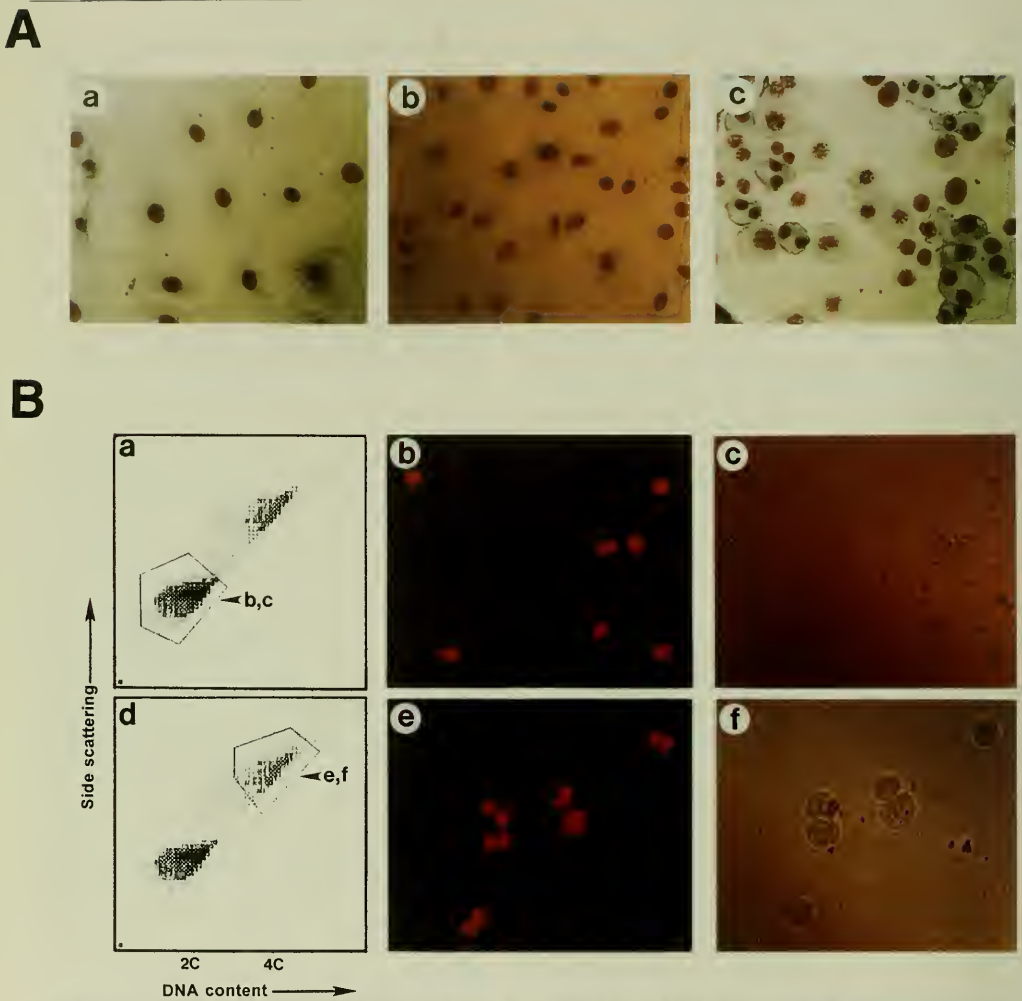


FIGURE 4. Microscopic and fluorescence-activated cell sorter (FACS) analysis of *2C*- and *4C*-arrested floating cells. (A) Cells were fixed with cold methanol, washed with phosphate-buffered saline (PBS), and then stained with Giemsa as recommended (Harleco). Mock-infected cells (a); murine leukemia virus-infected 3T3 cells (b); floating cells (c). Magnification 250X. (B) Approximately 1 to 2 x 10^6 floating cells were collected for FACS, washed once with PBS, then fixed in 70% ethanol for 30 minutes. Fixed cells were briefly pelleted and resuspended in PBS. Cells were then treated with RNase A (100 µg/ml) for 30 minutes at 37°C. Propidium iodide (10 µg/ml final concentration) was added to the cell suspension prior to the analysis. To remove the dead cells, the floating cells were subjected to Percoll gradient separation. Cells were gated with respect to side scattering and DNA-fluorescent intensity [panels (a) and (d)]. The sorted cells were sedimented onto a slide and examined at 100X magnification. Panels (b) and (e) show propidium iodide-stained *2C*- and *4C*-arrested floating cells, while panels (c) and (f) show the phase contrast of (b) and (e) preparations, respectively.

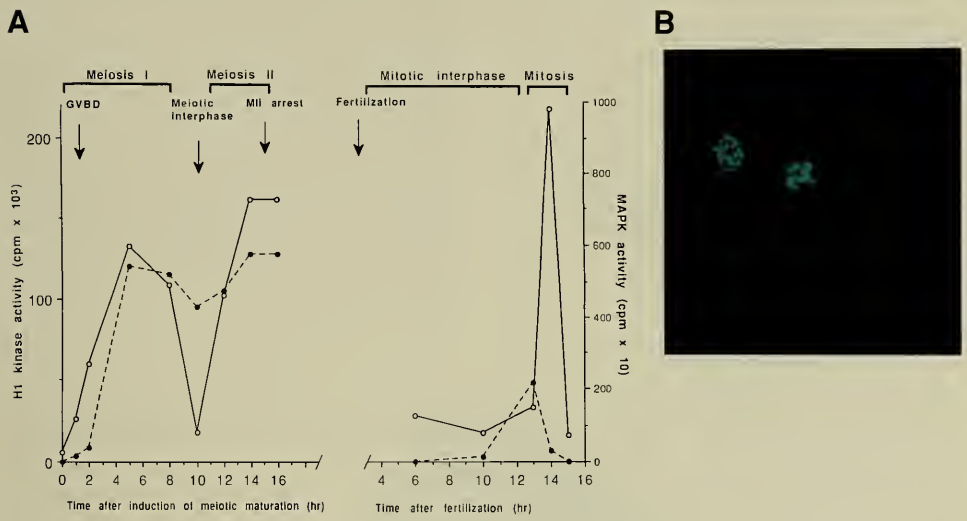


FIGURE 5. Histone H1 kinase and MAPK activities during meiosis and postmitotic cell cycle in mouse oocytes and embryos. Fully grown oocytes were collected from ovaries of 21- to 23-day-old B6C3 F_1 females 45 hours after stimulation with 7.5 IU of pregnant mare serum gonadotropin (PMSG). Oocytes complete with cumulus cells were used. For ovulated oocytes, females were injected with 7.5 IU PMSG followed 48 hours later with 7.5 IU human chorionic gonadotropin (hCG). Ovulated oocytes were collected from ampullar oviducts 14 to 16 hours after hCG injection. For in vitro fertilization, spermatozoa were collected from cauda epididymis of mature ICR males and incubated for 1 to 2 hours at 38.5°C to allow capacitation before insemination. The final concentration of spermatozoa was adjusted to approximately 2 x 10^5/ml. Oocytes and embryos were cultured in modified Whitten's medium supplemented with 4 mg/ml BSA at 38.5°C. Maturing oocytes were selected and pooled at three time points for synchronization: isolation from the follicle, at germinal vesicle breakdown (GVBD), and at first polar body extrusion. Under these conditions, synchronization is >90%. For the analysis of mitotically cycling embryos, successfully fertilized one-cell embryos were selected based on the presence of a second polar body and two pronuclei. The synchronization was >85% as determined by the time of nuclear breakdown and the first cleavage. (A) The histone H1 kinase assay for mouse oocytes and embryos was performed as follows: 10 oocytes or embryos were extracted in 10 µl of reaction buffer and the extracts were then mixed with 20 µl of reaction buffer containing 0.5 µg/ml histone H1 and [γ^{32}-P]ATP. The reaction was incubated at room temperature for 15 minutes and stopped by the addition of 30 µl of 2X sample buffer. Histone H1 kinase activity (—o—). For measuring MAPK activity in mouse oocytes and embryos, extracts from 50 oocytes or embryos were prepared in 1% SDS sample buffer and SDS-polyacrylamide gels were polymerized with 0.5 mg/ml MBP as substrate. After electrophoresis, SDS in gels was removed by incubation in 20% 2-propanol, 100 mM Tris-HCl (pH 8.0), and 5 mM β-mercaptoethanol; after denaturation-renaturation, a kinase assay was carried out in the polyacrylamide gel. The MBP phosphorylation activities were measured by autoradiography followed by scintillation counting of the corresponding bands (---●---). (B) Diamidino-phenylindole (DAPI) staining of oocytes at meiotic interphase. Oocytes were fixed with 3.7% formaldehyde and then stained with DAPI (magnification 100X).

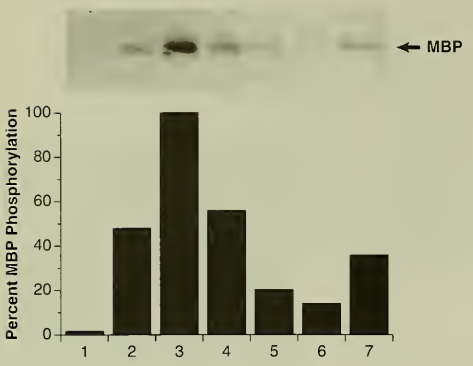


FIGURE 6. MAPK activation in cells expressing v-Mos. Cells maintained for 48 hours in 0.5% calf serum were washed twice in phosphate-buffered saline and then lysed in Triton X-100 buffer. The clarified lysate was subjected to immunoprecipitation with anti-MAPK antibody and protein A-agarose complex and tested for MAPK activity using myelin basic protein (MBP) as a substrate. Control 3T3 cells (lane 1); v-*raf*–transformed 3T3 cell line (BXB) (lane 2); floating cells (lane 3); v-*mos*–transformed cell lines, Tx3 (lane 4), Tx2 (lane 5), and Tx1 (lane 6); and c-*mos*–transformed cell line (PTS1) (lane 7). The v-*raf*–transformed constitutively activated MAPK was included as a control (lane 2). MBP phosphorylation was analyzed by SDS-PAGE on 18% gels, followed by autoradiography (upper panel). The kinase activity was also quantitated by scintillation counting and the data are presented in bar graphs as percentage of MBP phosphorylation (lower panel). The counts of the floaters set as 100% equaled 9350 dpm. Consistent with the higher specific activity of the c-*mos* product, PTS1 c-*mos*–transformed cells display high MAPK activity, but low levels of $p39^{mos}$ compared to the v-*mos*–transformed cells.

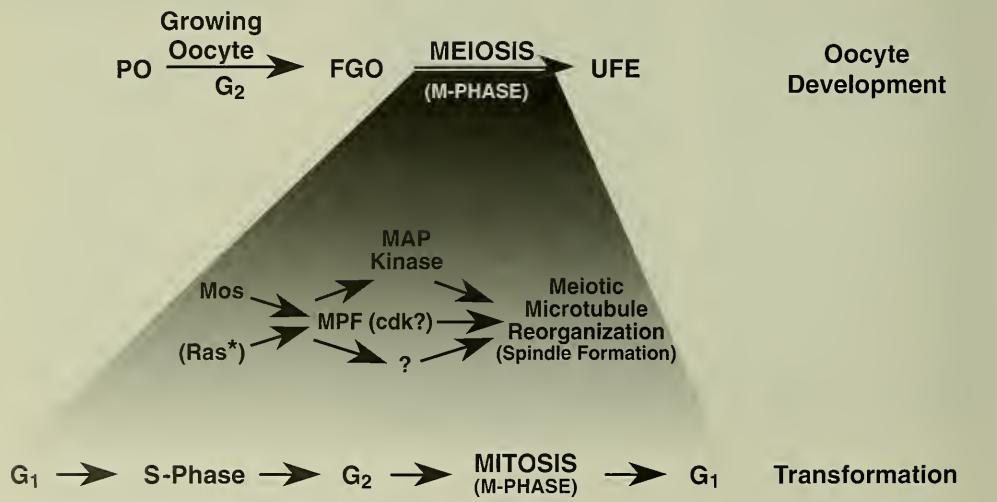


FIGURE 7. Mos in oocyte development and transformation. The Mos product is a regulator of meiotic maturation. The association of Mos with tubulin and MPF, as well as with MAPK activation, suggests that Mos contributes to the reorganization of microtubules that leads to spindle formation and the partitioning of chromosomes that occurs during meiosis I and II. The participation of Mos in other meiotic activities, such as nuclear envelope breakdown and chromosome condensation, is not excluded (as indicated by the question mark). The Ras oncoprotein can mimic the activity of Mos during oocyte maturation and in embryonic cleavage-arrest assays. It is suggested that the M-phase activity of Mos or Ras, inappropriately expressed during interphase of the somatic cell cycle, is responsible for the phenotype of transformed cells. Abbreviations include: PO, primordial oocyte; FGO, fully grown oocyte; UFE, unfertilized egg.

concurrently with the program of the somatic cell cycle (Figure 7). High-level Mos expression in somatic cells induces growth arrest, binucleation, and chromosome condensation. Our studies also show that Mos expression in somatic cells induces M-phase phosphorylation events at nonmitotic phases of the cell cycle (data not shown).

Since cells transformed by v-*mos* have been shown to be genetically unstable, binucleation can provide an explanation for the polyploidy observed in tumor cells. It is possible that Mos either directly or, more likely, indirectly (through initiating a cascade of kinases, including MAPK) interferes with the normal assembly of mitotic poles and asters and prevents proper positioning of the mitotic apparatus for cytokinesis. It is possible that inappropriate chromosome condensation and binucleation contribute to the genetic instability which is a component of tumor cell progression. The sister chromatid exchange and chromosome segregation that occurs uniquely during meiosis are likely to employ different checkpoint functions for fidelity than the checkpoints that control mitosis in somatic cells. Superimposing a meiotic program on the somatic mitotic cell cycle could markedly influence and compromise somatic cell checkpoint function. Other oncogenes, such as *ras*, *src*, and *raf*, that have M-phase activity or induce meiotic maturation also activate MAPK. Thus, morphological transformation may be a specific biological process induced by a meiotic M-phase program that runs concurrently with all stages of the somatic cell cycle program to generate many of the phenotypes of transformed cells (Figure 7).

The *met* Proto-Oncogene Mediates Mesenchymal to Epithelial Cell Conversion

Last year, we reported that the *met* proto-oncogene is involved in the formation of epithelial lumenal structures. We observed that in most epithelial organs Met is expressed at high levels in the cells that border lumenal structures (1992 Annual Report). Moreover, we found that treating certain epithelial carcinoma cell lines with the Met ligand, HGF/SF, induced lumen formation in vitro (1992 Annual Report). We also reported last year that tumorigenic activity of the *met* proto-oncogene in NIH/3T3 fibroblast cells occurs through an autocrine mechanism. While the mouse Met product (Met^{mu}) was activated, HGF/SF^{mu} expressed endogenously in these cells did not efficiently activate human Met (Met^{hu}). However, Met^{hu} was highly tumorigenic when coexpressed in NIH/3T3 cells with its human ligand, HGF/SF^{hu}.

Histopathological examination of the tumors generated by this autocrine mechanism revealed unexpected results. First, we observed that tumor explants of NIH/3T3 cells coexpressing Met^{hu} and HGF/SF^{hu} (HMH cells) displayed a lumen-like morphology in vitro and in vivo (data not shown). This appearance was quite unusual, since the lumenal morphology was thought to be specific for epithelial cells in culture and epithelial cell-derived carcinomas. Second, hematoxylin and eosin (H&E)-stained sections of the HMH tumors revealed areas with a trabecular pattern (Figure 8A), which is more characteristic of an epithelial adenocarcinoma than of the fibrosarcomas that typically develop from oncogene-transformed NIH/3T3 cells. In each of ten HMH tumors examined, we observed carcinoma-like focal areas that displayed typical epithelioid cell clusters (inset, Figure 8A). Third, examination of serial sections of the HMH tumors by transmission electron microscopy revealed that the tumor cells are connected by junctional complexes that contain desmosomes (arrows, Figure 8B, and upper right inset). The cells exhibiting desmosomes were derived from the HMH tumor, since they were positive for Met^{hu} expression (Figure 8B, upper left inset). The presence of desmosomes provides further evidence that the HMH tumor cells have converted to a carcinoma morphology.

Since desmosomes are the anchoring sites for cytokeratins, intermediate filaments that are epithelial cell specific, we tested HMH tumor cells in vitro for the expression of cytokeratins using anti-pancytokeratin antibodies (Figure 9A). While NIH/3T3 cells or another NIH/3T3 tumor cell line (MT) did not express cytokeratins (Figure 9A, panels 1 and 2, respectively), all cells from independently derived HMH tumors expressed high levels of cytokeratins (Figure 9A, panels 3 and 4, respectively). We also examined HMH and MT tumor sections for in vivo cytokeratin and vimentin expression. Vimentin is the mesenchymal-specific intermediate filament that is normally expressed by NIH/3T3 fibroblasts. As in the in vitro analyses, cytokeratins were highly expressed in vivo in the HMH tumor (Figure 9B, panel 1), but not in the MT control tumors (Figure 9B, panel 4). However, both HMH and MT tumors expressed high levels of vimentin (Figure 9B, panels 2 and 5, respectively). Staining of the HMH tumor with

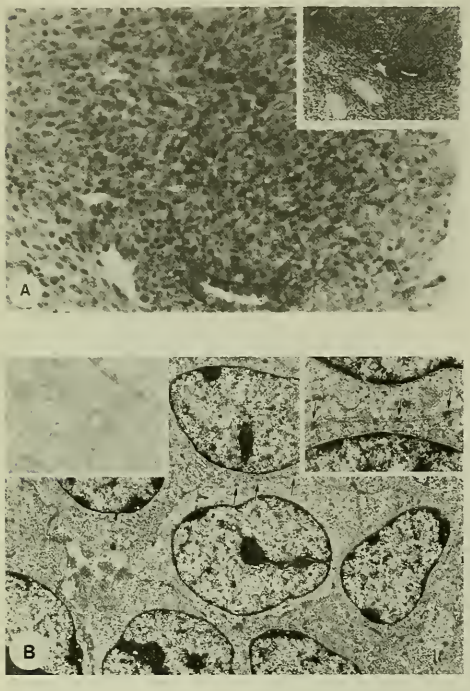


FIGURE 8. HMH tumors have carcinoma-like focal areas and desmosomal structures. (A) Light microscopic analysis of an H&E-stained HMH tumor section. This analysis reveals that the section contains large regions with carcinoma-like morphology. The cells in this region are polygonal and darkly stained and form islands of glandular cells with minimal connective tissues. These areas resemble a solid, poorly differentiated adenocarcinoma. The low magnification (inset, 160X) shows that a major portion of the tumor is carcinoma-like, but parts of the tumor retain the spindle-shaped fibrosarcoma pattern typically observed with NIH/3T3 cell tumors. The spindle-shaped sarcoma cells are seen in the lower left corner at high magnification (400X). (B) Transmission electron microscopic analysis of the carcinoma area of the HMH tumor section. A serial section of the same tumor shown in panel A analyzed by electron microscopy (magnification 11,300X) exhibits close cell-to-cell contact with desmosome-like junctions (see arrows and right inset at higher magnification, 18,500X). A serial section of the same tumor analyzed with C28 anti-Met^{hu} and immunogold electron microscopy labeling shows Met^{hu} staining in the cells (left inset, magnification 11,300X).

antibody specific for Met^{hu} (Figure 9B, panel 3) confirmed that this tumor arose from the HMH cells. The frequent formation of lumenal structures in the HMH sections is revealed in the Nomarski view (Figure 9B, panel 6). These analyses demonstrate that the HMH cells derived from NIH/3T3 fibroblast cells have authentic epithelial properties that can originate only from the coexpression of the Met^{hu} receptor and its ligand, HGF/SF^{hu}.

The acquisition of epithelial properties by the fibroblast-derived HMH cells mimics the mesenchymal to epithelial conversion of cells during kidney, ovary, and testis organogenesis. These differentiating cells are known to express both vimentin and cytokeratin during embryonic development, and Met is expressed at high levels in these cells (data not shown). These studies implicate the Met:HGF/SF signal transduction pathway in embryonic kidney development. The coexpression of Met, vimentin, and cytokeratin during the differentiation of mesenchymal to epithelial cells in the developing kidney parallels the conversion observed in the HMH tumors.

In adult tissues, the expression of different intermediate filaments is cell-type specific. The coexpression of cytokeratins and vimentin is ordinarily restricted to early development, with two notable exceptions: wound healing and certain types of neoplasia. Our results suggest that Met activation in mesenchymal cells by its ligand, HGF/SF, may play a similar role in the conversion of these cells to the epithelial cells of the wound.

It is interesting to note that giant cell lung carcinomas and epithelial lumenal cells in primary breast medullary duct and mucinous adenocarcinomas all express high levels of vimentin, as do epithelial portions of the kidney during tumor progression. These carcinomas may originate in a fashion similar to that of the carcinoma-like region of the HMH tumors. However, we cannot exclude that other growth factors and receptors may be required for or contribute to the changes necessary for the carcinoma-like conversion.

Met^{hu} Is Expressed in Fibroblast Cells and Is Overexpressed in Sarcomas

Mouse NIH/3T3 fibroblasts expressing Met^{hu} and HGF/SF^{hu} are highly tumorigenic in nude mice via an autocrine mechanism (1992 Annual Report). This result led us to examine human cell lines established from various

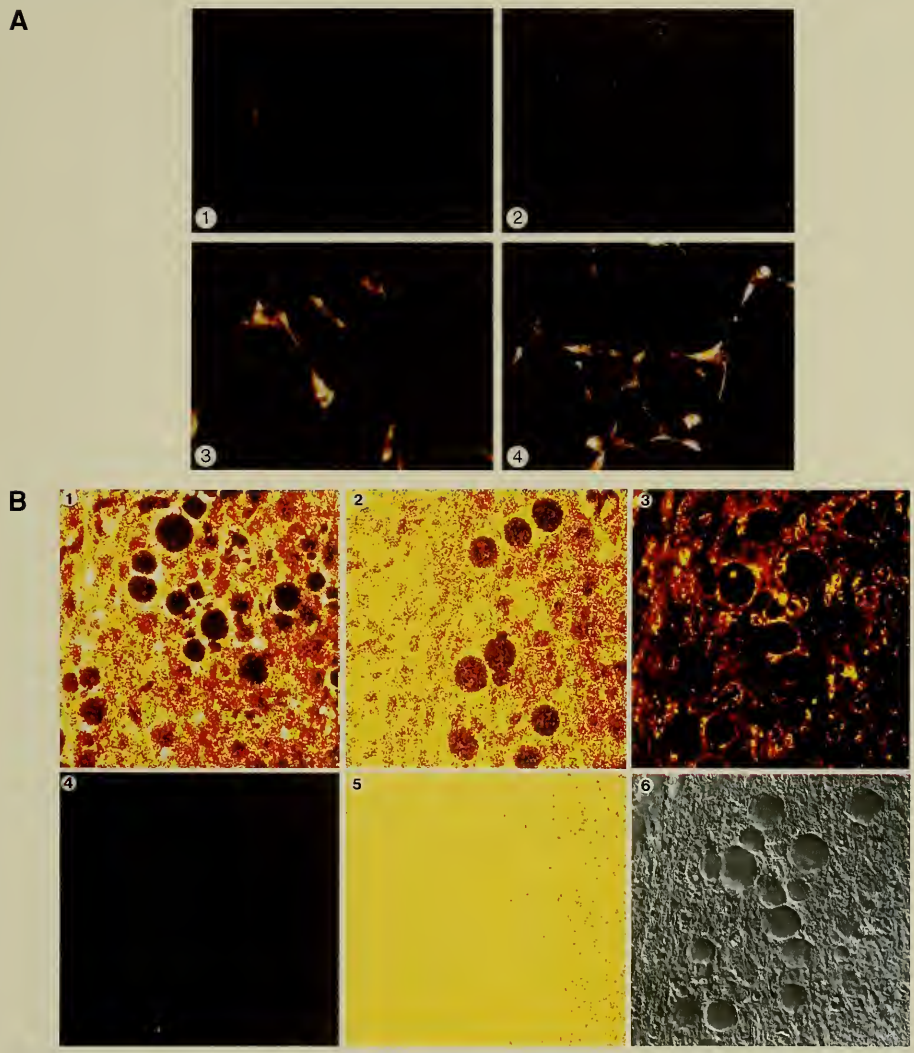


FIGURE 9. HMH tumor cells express cytokeratins in vivo and in vitro. (A) The same number of NIH/3T3, MT, and HMH cells were cultured in vitro and stained with cytokeratin. Panels: 1, NIH/3T3 cells; 2, MT cells; 3 and 4, independently derived HMH tumor cell explants (scale bar, 50 μm). (B) HMH tumors stained with antibodies against cytokeratin and vimentin. Paraffin sections of HMH or MT tumors were examined by confocal laser scan microscopy after staining with anti-pancytokeratin (panels 1 and 4) or anti-vimentin (panels 2 and 5) antibodies. MT tumors express vimentin (panel 5), but not cytokeratin (panel 4). A Nomarski view (panel 6) and Met^{hu} antibody staining (panel 3) are shown for the HMH tumor (magnification 560X).

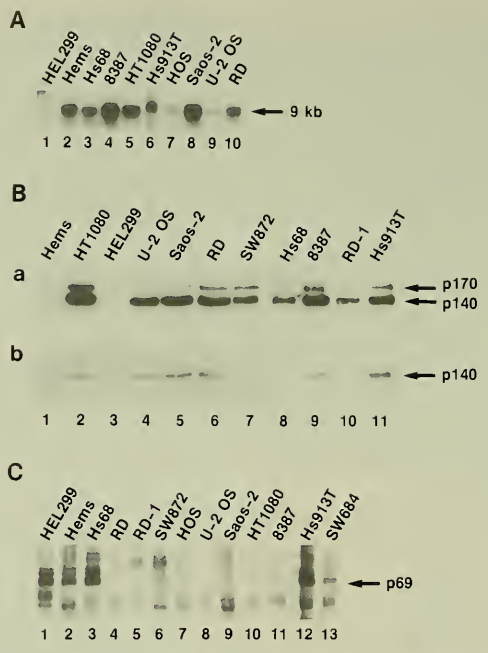


FIGURE 10. Met and HGF/SF expression in human cells. (A) Twenty micrograms of total RNA were loaded per lane for Northern analysis using full-length *met*^{hu} cDNA as a probe. Lane 1, HEL299; lane 2, Hems; lane 3, Hs68; lane 4, 8387; lane 5, HT1080; lane 6, Hs913T; lane 7, HOS; lane 8, Saos-2; lane 9, U-2 0S; lane 10, RD. (B) One milligram of cell lysate was immunoprecipitated with anti-C28 peptide antibody, followed by SDS-PAGE and immunoblotting with a Met monoclonal 19S antibody (a) or anti–P-Tyr antibody (b). Lane 1, Hems; lane 2, HT1080; lane 3, HEL299; lane 4, U-2 0S; lane 5, Saos-2; lane 6, RD; lane 7, SW872; lane 8, Hs68; lane 9, 8387; lane 10, RD-1; lane 11, Hs913T. (C) Cells were metabolically labeled with [^{35}S]methionine and [^{35}S]cysteine for 5 hours. One milliliter of supernatant was concentrated tenfold in a Centricon apparatus (10K cut-off); the volumes were adjusted to 0.35 ml with RIPA buffer for immunoprecipitation with HGF/SF monoclonal antibody A3.1.2. Lane 1, HEL299; lane 2, Hems; lane 3, Hs68; lane 4, RD; lane 5, RD-1; lane 6, SW872; lane 7, HOS; lane 8, U-2 0S; lane 9, Saos-2; lane 10, HT1080; lane 11, 8387; lane 12, Hs913T; lane 13, SW684.

sarcomas for Met and HGF/SF expression. In addition, we looked for Met expression in low-passage nonimmortalized human primary fibroblast cultures. Furthermore, we examined these cells for HGF/SF, which is known to be expressed in mesenchymal cells.

Total RNA extracted from nonimmortalized human fibroblast cell cultures and from human sarcoma cell lines was analyzed by Northern blot analysis for *met* RNA expression using a full-length *met* cDNA probe (Figure 10A). The major *met* mRNA 9-kilobase transcript was present in all of the samples tested (Figure 10A and Table 1). We also examined these cells and cell lines for Met protein expression by immunoblot analysis. Met was immunoprecipitated from cell lysates using anti-C28 peptide antibody directed against the C terminus of Met^{hu}, then immunoprecipitates were subjected to SDS-polyacrylamide gel electrophoresis (SDS-PAGE) followed by immunoblot analysis using 19S monoclonal antibody directed against the intracellular domain of Met^{hu} (Figure 10B). Low levels of p140^{Met} and its precursor p170^{Met} were detected in several of the primary fibroblast cultures (Figure 10Ba, lanes 1, 3, and 8; Table 1), but much higher levels were

TABLE 1. Met and HGF/SF expression in human fibroblast and sarcoma cell lines

	Met*	*met*†	HGF/SF‡	Scatter activity§		P-Tyr–Met‖
1. Human diploid fibroblast						
HEL 299 (fetal lung)	+	+	++++	4+	150	–
Hems (fetal muscle)	++	+++	+++	3+	52	+
Hs68 (newborn fibroblast)	++	++	++++	1+	6	–
Malme-3 (skin fibroblast)	+	ND¶	++++	1+	2	ND
2. Fibrosarcoma						
8387	++++	++++	++	1+	3	+++
HT1080	++++	+++	+	1+	6	+++
Hs913T	++++	++	++++	1+	3	++++
SW684	++++	ND	++	1–	0	+++
3. Osteogenic Sarcoma						
HOS	+++	+	–	1–	0	+
SAOS-2	+++	++++	+	1+	6	++
U-205	+++	+	–	1–	0	++
4. Chondrosarcoma						
SW1353	++	ND	–	ND		ND
5. Rhabdomyosarcoma						
RD	+++	++	–	1+	3	++
RD-1	++	+	+	1+	3	–
A204	–	ND	–	ND		–
A673	+	ND	–	1–	0	+
Hs729	+++	ND	+	1+	6	+++
6. Leiomyosarcoma						
SK-LMS-1	++++	ND	+	1+	6	++++
SK-UT-1B	+	ND	–	1+	6	+
7. Liposarcoma						
SW872	+++	+++	–	ND		–
8. Mesodermal tumor						
SK-UT-1	++++	ND	–	1–	0	+++
9. Synovial sarcoma						
SW982	+	ND	+++	1–	0	+
10. Melanoma						
Malme-3M	+++	ND	–	1–	0	+++
WM115	+	ND	+	1–	0	+
WM266-4	+	ND	–	1–	0	–

* Met protein level was assessed by immunoprecipitation and Western analysis.
† *met* gene expression was detected by Northern analysis.
‡ HGF/SF protein level was assessed by immunoprecipitation analysis.
§ Scatter activity was assayed with MDCK cells.
‖ P-Tyr–Met was analyzed by Western analysis with anti–P-Tyr.
¶ Not determined.

present in most of the human sarcoma cell lines tested (Figure 10Ba, lanes 2, 4 to 7, and 9 to 11; Table 1). These levels are reminiscent of the high levels of Met detected in NIH/3T3 cells transformed with *met* (data not shown). These analyses also show that there is no direct correlation between the levels of *met* mRNA (Figure 10A) and Met protein (Figure 10Ba). For example, the levels of *met* mRNA detected in 8387 and Saos-2 cells are higher than the levels detected in HT1080, while the highest level of protein is detected in HT1080. By contrast, the same levels of *met* RNA are detected by Northern analysis in Hems and HT1080, but the level of Met in the Hems fibroblast cell line is barely detectable (Figure 10Ba), while HT1080 is the highest. These results show that posttranscriptional mechanisms significantly regulate the steady state levels of Met protein expression. The high levels of Met in the sarcoma cell lines and its presence in fibroblast cells were novel findings, since Met expression in cells and tissues is preferentially restricted to epithelial cells (1988 Annual Report). By contrast, HGF/SF expression is usually restricted to mesenchymal cells.

We also determined the levels of immunoprecipitable HGF/SF in 5-hour–conditioned medium prepared from metabolically labeled fibroblast and sarcoma cell lines by SDS-PAGE analysis (Figure 10C). Abundant levels of the p69 subunit of HGF/SF were present in the three primary fibroblast cell cultures (Figure 10C, lanes 1 to 3; Table 1), but only one sarcoma cell line (Hs913T) showed a high level (Figure 10C, lane 12). The level of HGF/SF in the 72-hour culture was also determined by scatter assays on MDCK cells (Table 1). High levels of scatter activity were detected in the HEL299 and Hems diploid fibroblast cultures. However, the scatter activity in Hs68 and Hs913T cells did not correlate with the levels of HGF/SF product detected (Figure 10C), suggesting that the activity of Met or HGF/SF in these cells may be modified by some mutation. With the exception of Hs913T cells, the levels of HGF/SF and scatter activity were low in cells that expressed high levels of Met (Table 1). NIH/3T3 cells overexpressing Met^{mu} also show a marked reduction in HGF/SF expression, which is presumably due to Met receptor-dependent removal of the ligand, since Met^{mu} was readily detected with phosphotyrosine (anti–P-Tyr). We therefore examined Met^{hu} from the human fibroblast cell cultures and sarcoma cell lines by reprobing the immunoblot (Figure 10Ba) with anti–P-Tyr (Figure 10Bb). Met was weakly reactive with anti–P-Tyr in the primary fibroblast Hems cell line (Figure 10Bb, lane 1), suggesting that Met may be activated by HGF/SF in an autocrine fashion. In contrast, these analyses showed that, in general, where Met is overexpressed, it is highly reactive with anti–P-Tyr antibody (Figure 10Bb, lanes 2, 4 to 6, 9, and 11; Table 1), especially in the fibrosarcoma cell lines. These results show striking parallels with the results obtained with NIH/3T3 cells transformed by and overexpressing Met.

Cells that synthesize growth factor and express the cognate receptor have the potential for either internal or external autocrine-mediated growth. Met and its ligand, HGF/SF, were first shown to interact in a paracrine fashion; HGF/SF is produced by cells of mesenchymal origin such as fibroblasts and smooth muscle cells and causes mitogenicity, cell movement, and morphogenesis of epithelial cells that express the Met receptor. In addition to this paracrine model, we have shown that signaling can occur in an autocrine manner in NIH/3T3 cells that express both HGF/SF and Met. Here we show that some primary fibroblast cells in culture produce abundant HGF/SF, and express low levels of Met protein, suggesting that an autocrine loop of Met:HGF/SF is utilized in these cells.

Met:HGF/SF autocrine interaction may be a fundamental property of fibroblast cells in culture as well as in other mesenchymal cells in vivo. In early studies, *met* RNA transcripts were detected in primary human foreskin fibroblast cells (1986 Annual Report). It is possible that this autocrine loop could contribute to both the mitogenic and morphogenic phenotypes of these cells. Spontaneous transformants of NIH/3T3 fibroblasts were found to have the *met* proto-oncogene amplified and overexpressed, and Met overexpression in NIH/3T3 fibroblasts was found to induce fibrosarcomas via an autocrine mechanism. The fact that Met is overexpressed in most of the human sarcoma cell lines tested suggests that Met may be involved in the development of soft tissue sarcomas.

Gene Expression in Eukaryotes Section

Stephen H. Hughes

Stephen V. Angeloni
Eugene V. Barsov
Paul L. Boyer
Mark J. Federspiel
Andrea L. Ferris
Iris Givol
Roberta J. Greenhouse
Amnon Hizi
Dolores P. Lana
Valerian Pinto
Lisa S. Shelton
Pramod Sutrave
Leanne Tesene
Jeannette M. Whitcomb

In 1963, Howard Temin published a paper in which he proposed that Rous sarcoma virus (RSV) replicates via a DNA intermediate (Temin, Virology 20:577, 1963). Since RSV has an RNA genome and, at that time, there was no direct evidence that RNA could ever be copied into DNA, Temin's hypothesis was not widely accepted. It was not until 1970, when both Temin and David Baltimore demonstrated that retroviral particles contain the enzyme reverse transcriptase (RT) and showed that this enzyme could actually copy RNA into DNA, that Temin's idea was generally accepted. Because the properties of RT were unexpected, it has been the subject of intense investigation from the moment it was discovered. The realization that the disease AIDS is caused by a retrovirus, human immunodeficiency virus type I (HIV-I), has focused research on retroviruses in general and on RT in particular. The available anti-AIDS drugs (AZT and ddI, for example) are RT inhibitors. Like many other retrovirologists, we have a substantial interest in HIV-I and have focused primarily on RT both because it is a target for drug development and because it is a complex and fascinating enzyme. We believe that a better understanding of the structure and function of RT should facilitate the development of more effective anti-AIDS drugs.

Retrovirus research had its origins with observations, made early in this century, that these viruses could cause tumors in animals. In most cases, this oncogenic potential derives from the integration of a DNA copy of the retroviral genome into the host cell genome. On rare occasion, the retroviral genome is inserted near a cellular oncogene which activates the oncogene and leads to the oncogenic transformation of the host cell. In some cases, this insertion also leads to the capture of the cellular oncogene by the retrovirus. Retroviruses that have captured cellular oncogenes are highly oncogenic; these retroviruses provided both the intellectual foundation and the research tools that led to the discovery of cellular oncogenes.

These highly oncogenic viruses are naturally occurring vectors and provided the inspiration, and in some cases the actual models, for the creation of retroviral vectors in the laboratory. Most of the naturally occurring vectors have lost part, or all, of one or more of their genes and require a helper virus for their propagation and, as might be expected, most of the retroviral vectors that have been created in the laboratory are replication defective. RSV is exceptional among the naturally occurring vectors in that it has acquired the cellular oncogene *src*, but still retains a full complement of viral genes and remains replication competent. We reasoned that it should be possible to insert a variety of other genes into a cloned DNA copy of the RSV genome in place of *src* and, in so doing, produce retroviral vectors that have the advantage of being replication competent. We have, over the past decade, developed a set of retroviral vectors that are now in wide use in laboratories throughout the world. One of the disadvantages of this system is that RSV itself, and the vectors that derive from it, do not efficiently infect mammalian cells, primarily because mammalian cells lack a specific receptor that the virus needs to enter the cell. Recently, Paul Bates and John Young, working in the laboratory of Harold Varmus (University of California at San Francisco), cloned the subgroup A receptor and showed that the expression of this

receptor in mammalian cells allows the cells to be efficiently infected by subgroup A avian retroviruses. We are collaborating with this group and have used a cDNA clone that encodes the subgroup A receptor to create transgenic mice that express the receptor. These mice can be specifically and efficiently infected with subgroup A vectors.

Ski is one of the many cellular oncogenes originally discovered as the transforming principle in an oncogenic avian retrovirus. Although *ski* is a relatively weak oncogene, it is particularly interesting because, in addition to being able to induce the proliferation of cultured chick embryo fibroblasts (CEFs), it has myogenic potential, both in cultured cells and in transgenic animals. We are attempting to discover how *ski* works. The Ski protein is localized in the nucleus and a growing body of evidence suggests that, either directly or indirectly, Ski can affect the transcription from certain muscle-specific promoters.

HIV-1 RT

HIV-1 RT has a central role in the life cycle of the AIDS virus: it is the enzyme responsible for copying the single-stranded RNA genome found in virions into the double-stranded linear DNA form which is subsequently inserted into the genome of the infected cell. Given the importance of RT, it is reasonable to expect that inhibitors of RT would be potent antiviral compounds; in fact, the great majority of anti-AIDS drugs that have been discovered so far are inhibitors of RT. The available RT inhibitors can be grouped into two broad classes: nucleoside inhibitors and non-nucleoside inhibitors. The nucleoside inhibitors are analogs of the precursors that are used by RT to make the DNA copy of the HIV-1 genome. Most of the nucleoside inhibitors, including the drugs that are being used to treat AIDS clinically (e.g., ddC, ddI, and AZT), are incorporated into DNA by RT and, when incorporated, block the subsequent growth of the DNA strand. By contrast, the nonnucleoside inhibitors bind directly to RT and block its enzymatic activity. The precise way in which the nonnucleoside inhibitors work is unclear; however, it is possible that these compounds interfere with the flexibility of the enzyme, a flexibility that is essential for activity.

Unfortunately, HIV-1 is a highly variable virus. The virus uses variation to escape the host's immune system and to evade all of the known inhibitors of RT. Although all of the RT inhibitors can initially block the replication of the virus, after a short time resistant viruses emerge. As expected, the mutant viruses contain altered forms of RT that are insensitive to the specific drug that was used against the virus. Since the nucleoside and nonnucleoside inhibitors act in different ways, different mutations in RT confer resistance to the nucleoside and nonnucleoside inhibitors.

These observations suggest that it may be possible to treat patients simultaneously with a combination of drugs — in particular, with combinations that give rise to different resistance mutations. The idea behind this type of therapy is that the virus will be forced to make multiple changes in response to two (or more) simultaneous challenges. It has been suggested that it may even be possible to constrain RT such that, if the virus attempts to make all the changes simultaneously, the enzyme will be unable to function properly. While this may be possible, the great difficulty is that the virus has been able to find ways to change RT structure to avoid each of the known drugs. Put another way, all of the available drugs are targeted to portions of RT that the virus can afford to change. Ideally, we would like to be able to discover or to create drugs that would interact with positions of the enzyme that are essential for its function, portions that are so critical that, if the virus tried to change them, RT would not be able to function properly.

Considerable progress has been made in defining essential parts of HIV-1 RT. We have used site-directed mutagenesis to identify positions in RT where even the most conservative change destroys the ability of the enzyme to copy RNA into DNA. Even more important, the structure of the enzyme is known in considerable detail. This structure, together with data from the site-directed mutagenesis, identifies the structure of those portions of RT critically important to its function. Knowing the structure of these critical areas affords the opportunity to design drugs to bind to these key elements, or perhaps more simply, to modify some of the available drugs so that they will interact more strongly and specifically with the key elements of RT.

Two groups (those of Thomas Steitz at Yale University and Edward Arnold at Rutgers University) have made important contributions to our current understanding of the structure of HIV-1 RT. Dr. Steitz's group has published a structure of a complex between HIV-1 and the

nonnucleoside inhibitor nevirapine at 3.5 Å resolution. Using HIV-1 RT and an Fab fragment of a monoclonal antibody, both of which we supplied, Dr. Arnold and his colleagues have grown crystals that diffract to high resolution. We believe that the Fab fragment stabilizes the structure of RT, which allows the growth of better crystals. Dr. Arnold and his colleagues have been able to prepare crystals with double-stranded synthetic DNA specifically bound to HIV-1 RT in a ternary complex with the Fab fragment. The structure of this DNA-containing complex has been solved to 3.0 Å resolution.

HIV-1 RT is synthesized as part of a large polyprotein precursor and is released by cleavage with a virally encoded protease. In the virion, HIV-1 RT is a heterodimer composed of two subunits, p66 and p51. The larger subunit, p66, is 560 amino acids in length and contains both the polymerase and RNase H domains. p51 contains the first 440 amino acids of the p66 subunit, which corresponds to the polymerase domain. The additional 120 amino acids of p66 comprise the RNase H domain. The polymerase domains of p66 and p51 each contain the same four subdomains, which have been called fingers, palm, thumb, and connection. Although the folding of the polypeptide chains within these four subdomains is similar in p66 and p51, the relative arrangement of the subdomains is significantly different. As a consequence, the p66/p51 heterodimer has only one DNA-binding groove and one polymerase active site, which is in the palm subdomain of p66. The corresponding region of p51 is buried in the structure and does not come in contact with nucleic acid. The asymmetry of the two subunits of the p66/p51 heterodimer also means that any point mutation in the region of the viral genome that encodes the polymerase domain causes changes at two different positions in the HIV-1 RT heterodimer.

The DNA is bound in a large groove on the surface of HIV-1 RT. The double-stranded DNA segment used in most of the crystallography experiments is a 19-base/18-base template primer that corresponds in sequence to the primer binding site in the genome of HIV-1. The 3′ end of the primer strand is in close contact with the active site for polymerization. There is, in the expected position at the 3′ end of the primer strand, a specific binding site for mercurated UTP, which unambiguously identifies the nucleoside triphosphate-binding site. The one-base overhang was deliberately designed to accept an incoming T so that it would be possible to bind AZT triphosphate (or analogs of this molecule) to the active site of the polymerase. The orientation of RNase H places the active site of this domain close to the bound nucleic acid.

To complement these structural studies, we are extending the genetic analysis of HIV-1 RT we began several years ago. As discussed in the 1992 Annual Report, we began this analysis by making insertion and deletion mutants; however, a thorough genetic analysis requires specific point mutants. Because we wanted to analyze a large number of HIV-1 RT mutants, we felt that it was essential to develop rapid methods to create and evaluate specific point mutants. We made a series of 16 *Bsp*MI cassettes that make it possible to change any amino acid (or group of amino acids) in HIV-1 RT. The cassettes take advantage of the properties of the restriction endonuclease *Bsp*MI, which cleaves DNA a discrete distance outside its recognition sequence and leaves a four-base protruding end. The HIV-1 RT cassettes contain two *Bsp*MI recognition sequences in opposite orientations. After *Bsp*MI digestion, a synthetic DNA fragment encoding the desired mutation can be ligated into the cassette. All of the mutants have been verified by DNA sequencing.

Each of the mutants is tested for RNA-dependent DNA polymerase activity, which can be done in crude bacterial extracts, since laboratory strains of *Escherichia coli* do not contain significant amounts of this activity. The mutant enzymes are also tested for DNA-dependent DNA polymerase and RNase H activity. Since *E. coli* extracts do contain large amounts of endogenous DNA-dependent DNA polymerase and RNase H, we have used in situ gel assays that allow us to separate HIV-1 RT from the *E. coli* enzymes with similar functions. The in situ assays have the advantage that a group of mutants can be simultaneously evaluated on a single gel. We have prepared more than 250 different site-directed mutants using the cassettes and have made a preliminary evaluation of their properties.

The purpose of this analysis is to define which positions in RT can, and which cannot, be mutated without significantly affecting the activity of either the polymerase or the RNase H. Having a detailed crystallographic structure for HIV-1 RT simplifies both the planning of the mutational analysis and the interpretation of the data.

We have found a number of positions where HIV-1 RT is quite sensitive to mutation. For example, mutating amino acids at the polymerase active site dramatically affects polymerase activity. However, this analysis has also yielded some unexpected results. We have found that there are positions in the polymerase domain where mutations selectively inactivate RNase H, but not polymerase. Examination of the structure suggests that these mutations cause an alteration in the positions of the nucleic acid, and, as a consequence, prevent the nucleic acid from making proper contact with the active site of RNase H.

We have also constructed mutants that correspond to the known drug-resistant variants. We are trying to understand precisely how both the nucleoside-resistant mutants and the nonnucleoside-resistant mutants work. There are good reasons to believe that most (or all) of the mutants that are resistant to the nonnucleoside inhibitors are resistant because the nonnucleoside inhibitors bind less effectively to the mutant RTs. One of the interesting conclusions from our analysis is a clear demonstration that, although the known nonnucleoside inhibitors all seem to bind to the hydrophobic pocket that lies under the active site in the 66-kilodalton subunit of HIV-1 RT, individual inhibitors have specific interactions with overlapping, but not identical, sets of amino acids that line this hydrophobic pocket. The mechanism by which the mutants confer resistance to the nucleoside inhibitors appears to be more subtle; most of these mutations are not located at the active site for polymerization where the nucleoside inhibitors are incorporated (see Figure 1).

As part of our collaboration with Dr. Arnold's group, we are attempting to understand precisely how individual mutations confer resistance to the nucleoside inhibitors. We know from mutational analysis that RT does not usually tolerate mutations at the polymerase active site; presumably, this is the reason why most of the resistance mutations occur elsewhere. However, it does raise an important question: how do mutations that are not in the immediate vicinity of the polymerase active site affect the enzyme's ability to select (or reject) a particular nucleoside analog? The first hints came from a careful examination of the RT structure. Although most of the mutations that cause resistance to nucleoside inhibitors are not at the polymerase active site, several of these mutations are in positions that appear to be able to make contact with either the single-stranded region of the nucleic acid template or with the double-stranded template/primer. These observations suggest that the mutations that confer resistance to the nucleoside inhibitors alter the position of the template-primer and, in so doing, alter the structure of the nucleic acid-protein complex at the polymerase active site. We have done biochemical experiments that were designed to test this hypothesis. We have shown that two mutations that confer resistance to nucleoside inhibitors (Glu89Gly and Leu74Val) alter the position of the template primer.

Retroviral Vectors and Cellular Receptors

Retroviral vectors have been used extensively to deliver and express genes in cultured cells and in animals. Most retroviral vectors are defective, which, depending on the experiment, can be either an advantage or a disadvantage. Since a defective virus does not usually spread efficiently even in the presence of a helper virus, most defective vectors carry selectable markers. This approach is not usually helpful in vivo. A solution that has been useful in vivo is to develop replication-competent vectors. We have developed a series of replication-competent avian retroviral vectors that have been used by a number of laboratories for a variety of in vivo (and in vitro) experiments. Despite the obvious advantages of replication-competent vectors, there are limitations. For example, the size of the foreign DNA that can be inserted into the vector is reduced, and the most efficient vectors of this type have been confined to the avian system. Moreover, the tendency of the replication-competent vectors to spread in vivo is itself both an advantage and a disadvantage. While there are several methods for controlling, or partially controlling, the spread and/or expression of replication-competent vectors in vivo, none of these methods are ideal. It is possible to introduce the recombinant virus to a specific portion of the animal (or developing embryo). Although the virus does spread, it has been possible to obtain expression (at least for a limited time) in particular regions of a developing chicken embryo (Morgan et al., Nature 358:188, 1992).

We have developed replication-competent vectors that express an inserted gene from an internal promoter and have shown that the internal promoter retains its appropriate tissue specificity in vivo, even in the context of the vector. It is also possible to control the spread of the vector

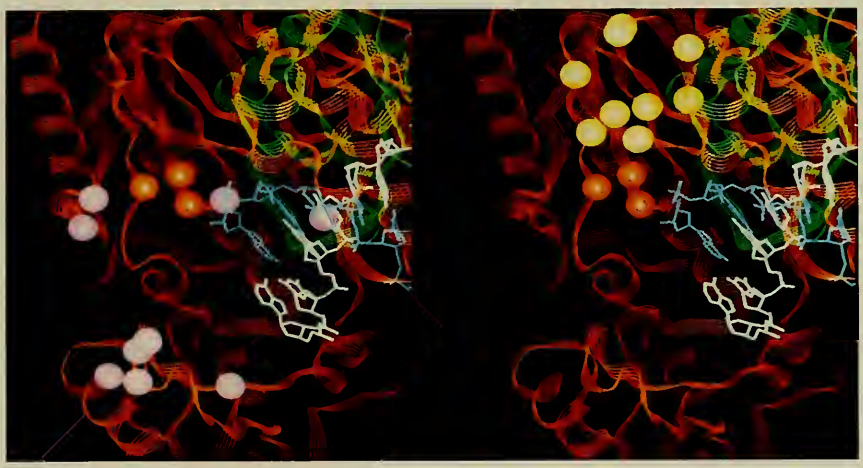


FIGURE 1. Mutations that cause resistance to both nucleoside inhibitors (left panel) and nonnucleoside inhibitors (right panel) are shown on a ribbon backbone of HIV-1 RT (red and green), complexed with double-stranded DNA (light gray and blue). The positions of the nucleoside resistance mutations are shown as pink spheres; the positions where nonnucleoside resistance mutations occur are shown as yellow spheres. The three red spheres are the aspartic acid residues that comprise the polymerase active site (amino acids 110, 185, and 186). (Figure courtesy of Dr. Edward Arnold.)

by making surgical chimeras between embryos that do and embryos that do not express a functional receptor for the viral vector. The vector will then be confined to whatever portion of the chimera derives from the susceptible embryo. This procedure, while effective, is quite labor intensive, since each experiment requires the construction of appropriate surgical chimeras. We have developed a method for genetically controlling the expression of the receptor; in so doing, we have made it possible to use our avian vectors in mice.

cDNAs that encode a biologically active receptor for the subgroup A avian leukosis viruses (ALVs) have been cloned. Mammalian cells that express this receptor can be efficiently infected by subgroup A ALVs and by vectors based on these ALVs (Young et al., J Virol 67:1811, 1993). There are additional blocks to the life cycle of ALVs in mammalian cells; for example, viral structural proteins are inefficiently expressed in mammalian cells and, in addition, assembly of the virion is partially or completely blocked. This means that, although mammalian cells that express the subgroup A receptor can be efficiently infected by subgroup A ALVs, the infected cells do not produce infectious virions, and subgroup A ALV vectors that are replication competent in avian cells are replication defective in mammalian cells, whether or not the mammalian cells express the receptor. This combination of systems has distinct advantages: the replication-competent ALV vectors can be grown to high titers on avian cells, and these high-titer viral stocks can be used to infect mammalian cells where the vectors are replication defective. We routinely derive our vectors by transfecting EV-O CEFs; these cells do not contain any endogenous retroviruses that are closely related to the ALV vectors. Not only are the cells used to derive the vector stocks free of closely related endogenous viruses, but the mammalian cells that are used as a target for the vectors do not contain endogenous viruses that are closely related to the ALV vectors. Since we do not use a helper cell line to derive the ALV vectors and, at every stage of the experiment, infect cells that lack closely related endogenous viruses, we can avoid the issue of viral recombination that has been a continual problem in the development and use of murine retroviral vectors.

The cloned subgroup A receptor can also be used to considerable advantage in vivo. We have developed lines of transgenic mice that carry and express this receptor. Not only do these animals give us the opportunity to extend the well-characterized and efficient vector system that we have developed to mice, it also gives us the opportunity to control the expression of the subgroup A receptor itself. We have linked the cloned receptor DNA to a muscle-specific promoter (the α_{sk}-actin promoter) and have used this segment to create several lines of transgenic mice. The mice express the receptor efficiently in skeletal muscle. Our previous experiments with the α_{sk}-actin promoter linked to the chloramphenicol acetyltransferase (CAT) reporter gene suggested that this promoter can also be active in the heart; several of the lines of α_{sk}-actin/A receptor mice do show expression in the heart. We were surprised to find that some of the mouse lines also express the transgene in smooth muscle, suggesting that something in the DNA sequence encoding the receptor somehow broadens the specificity of the α_{sk}-actin promoter. The other tissues that we have tested appear to express the transgene at lower levels.

Not only is the subgroup A receptor expressed in the muscles of the transgenic mice, it also confers on these mice susceptibility to infection by subgroup A ALV (see Figure 2). Siblings that do not carry the transgene cannot be infected by the vectors. As expected, the virus is confined to the muscle of the particular leg that is injected; there is no spread of the virus to the contralateral leg. This system provides us with a simple and effective way to induce the expression of a wide variety of genes in the muscles of these transgenic mice. We are developing additional lines of transgenic mice that will express the receptor in other tissues; for example, we have linked the subgroup A receptor to the β-actin promoter. This promoter is expressed in a variety of cells and tissues, which should give us access to most of the tissues in the animal. It should be relatively simple to produce additional lines of mice that have the receptor linked to, and under the control of, other cell-specific and muscle-specific promoters.

We believe that these mice have the potential to be used in several types of experiments; for example, since the RSV-derived vectors do not replicate in mammalian cells, this system should be quite valuable for cell lineage analysis.

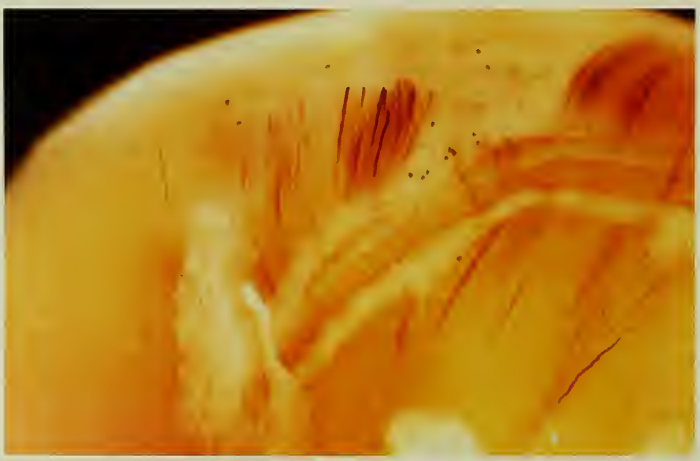


FIGURE 2. Thigh muscle of a transgenic mouse stained for human alkaline phosphatase (HAP). The leg muscle of a mouse carrying the α_{sk}-actin/subgroup A receptor transgene was injected with 10^3 to 10^4 infectious units of an RCASBP-HAP virus at 5 days of age. At 10 days of age, the mouse was killed and the injected leg was stained for HAP. Cells infected with the RCASBP-HAP virus are stained dark purple. The purple streaks are mature muscle fibers that express HAP; the round purple dots are satellite cells that express HAP.

In addition, because the avian retroviral vectors we have developed can easily be grown to high titers in cultured chicken cells, and because we have developed lines of transgenic mice that express the receptor in a defined cell type or tissue, the combination of the transgenic mice and the vectors should provide a powerful model system for gene therapy experiments. This system will complement recent progress in mouse genetics that has provided, and will continue to provide, useful models for many human genetic disorders. The combination of the transgenic receptor mice and the avian retroviral vectors we have developed should also be quite useful for investigations of the roles of particular genes in development and oncogenesis. One of the primary reasons we developed this system is to facilitate our investigations of the effects of *ski* expression on muscle growth and development.

Ski

We initially chose to study *ski* because it has two distinct effects in cultured cells: it induces CEFs to proliferate more rapidly and grow in soft agar (Stavnezer et al., J Virol 39:920, 1981). In addition, *ski* can induce cells isolated from the body wall of quail embryos to differentiate into muscle (Colmenares and Stavnezer, Cell 59:293, 1989). These two properties are not only distinct, they also appear, at least superficially, to be contradictory. Under most circumstances, increased proliferation does not lead to differentiation, nor is differentiation ordinarily associated with proliferation.

When we began our investigations, the only form of *ski* that had been characterized was the form present in the original retroviral isolates (v-*ski*) (Stavnezer et al., J Virol 57:1073, 1986). It was unclear whether v-*ski* was derived from one gene or two. It was possible that v-*ski* was derived from a gene fusion, and *ski*'s ability to induce both proliferation and muscle differentiation derived from the two unrelated antecedent genes. Dr. Stavnezer and his colleagues, who isolated the original v-*ski* viruses, had begun an analysis of the homologous chicken gene (c-*ski*). Ordinarily, these studies would have resolved the issue; however, their cloning efforts yielded two sets of genomic DNA clones, which they were not able to connect to form a single segment. One set of genomic clones contained exons present in the 5′ end of v-*ski*; the second set of genomic clones contained exons present in the 3′ end of v-*ski*. Although more than 65 kilobases of genomic sequences were isolated, the cloning efforts did not yield a contiguous set of sequences. Were there two genes, or one gene with large introns?

To answer this question, we isolated c-*ski* cDNA clones from chicken embryo body wall mRNA. These clones contained both the 5′ and 3′ sequences from v-*ski*, which proved that there was a single c-*ski* gene that must contain one or two relatively large introns. Although the cDNA analysis showed that v-*ski* was derived from a single gene, the possibility remained that c-*ski* was some sort of ancient gene fusion and that the ability of *ski* to induce both differentiation and proliferation was present in distinct domains. Subsequent analysis has shown that this idea is incorrect and that the mitogenic and myogenic effects of *ski* are not readily separable. In addition, the cDNA clones revealed something we had not anticipated: chicken *ski* mRNAs are alternatively spliced. Three distinct types of c-*ski* cDNAs were found, suggesting the presence of three types of c-*ski* message and, by implication, three distinct c-Ski proteins. The longest cDNAs appear to derive from mRNAs that contain all eight known coding exons. This form encodes a 90-kilodalton protein (p90). The other two forms lack the second coding exon and the sixth coding exon, respectively. Omission of the second exon does not alter the reading frame; however, omission of the sixth exon does change the reading frame, creating a termination codon at the boundary between the fifth and seventh exons. These cDNAs encode proteins of 85 and 60 kilodaltons, respectively.

It is often possible to infer something about the activity of a gene based on its homology with genes that are better understood. *Ski* has a slight homology to *myc*. This is intriguing, since *Myo*D and *myc* are both members of a superfamily of transcription factors that is defined by a helix-loop-helix motif. Despite the fact that the region of *myc* that has homology to *ski* lies in the portion of *myc* that contains the helix-loop-helix motif, the homology between *myc* and *ski* does not include the amino acids that define the helix-loop-helix motif. The only known gene that has significant homology to *ski* is *sno* (*ski*-related novel gene), which was discovered in a screen for human c-*ski* cDNAs (Nomura et al., Nucleic Acids Res 18:5489, 1989).

We introduced the various forms of c-*ski* and c-*sno* into retroviral vectors and used these vectors to study the biochemical and biological properties of these proteins. The biologically active forms of v-Ski, c-Ski, and c-Sno proteins are localized primarily in the nucleus. When the p85 and p90 forms of Ski protein are overexpressed, most of these proteins are found in small clumps in the nucleus, while p60 is dispersed somewhat more uniformly throughout the nucleus. We have used immunoelectron microscopy to study the structure of the Ski-containing clumps. The particles, which we call Ski bodies, are globular and are 0.15 μm to 2 μm in diameter. Some of the larger Ski bodies have an electron lucent core that does not seem to contain significant amounts of Ski protein. Since these observations were made in cells in which the c-Ski proteins were overexpressed, it is unclear whether the apparent differences in subnuclear localization are relevant. It should be remembered that all of the Ski proteins have, as far as we can measure, similar biological activities.

Both the *ski* and *sno* oncogenes are expressed at low levels in most tissues during normal development, so the pattern of expression provided little useful information about their normal function(s). We decided to test the function of these genes by making transgenic mice that overexpress c-*ski*. In the initial experiments, a truncated chicken c-*ski* cDNA (Δ29) was placed under the transcriptional control of the long terminal repeat (LTR) of murine sarcoma virus (MSV). Of the 44 founder mice, 3 showed a distinctive muscular phenotype. Independent lines were derived from these three animals and from a fourth founder animal that carried intact copies of the transgene but did not show the muscular phenotype. In addition to having substantially more muscle, mice from the three phenotypically positive lines have considerably less fat than control animals. We have made transgenic mice that carry a c-*sno* transgene; however, the *sno* transgenic mice we have made so far do not show this dramatic muscular phenotype.

The three phenotypically positive lines express high levels of RNA from the transgene in skeletal muscles, and we have recently found that there is significant expression in the long bones. There is relatively little or no mRNA in the other tissues we have tested. We do not as yet have a satisfactory explanation for this observation. We examined the skeletal muscles in more detail to determine whether the increase in muscle mass was due to hypertrophy or to hyperplasia. Most of the experiments that we have done so far have been carried out with mice from line 8566, which was chosen at random from the three lines with the muscular phenotype. Some of the muscles in this line have more than twice the cross-sectional area of the corresponding muscles in normal mice. This increase is due exclusively to hypertrophy.

Mammalian muscles are composed of several different fiber types that can be distinguished because they express different myosin heavy chain genes. In the *ski* transgenic lines that have the muscular phenotype, some but not all of the fibers undergo hypertrophy. We examined the hypertrophied muscles from line 8566 with antibodies that can distinguish the various myosin heavy chain isotypes and showed that type I, type IIa, and type IIx fibers did not undergo hypertrophy; however, type IIb fibers were hypertrophied. When the amount of c-Ski protein in the nuclei of the muscle fibers was measured by immunofluorescence, all of the fibers in muscles that were hypertrophied had approximately equal amounts of c-Ski protein, whether or not the individual fibers were hypertrophied. We cannot yet explain why some fiber types undergo hypertrophy while others do not.

To test whether the truncated form of Ski was important in the generation of the muscular phenotype, we linked the MSV LTR to two full-length c-*ski* cDNAs, 29 and 27. These transgenes are approximately as effective as the MSVΔ29 transgene in producing muscular animals. We also found that the α_{sk}-actin promoter, which we had previously shown to be muscle specific in transgenic mice, was not able to induce the full muscular phenotype if it was linked to any of the cDNAs (Δ29, 27, or 29) that were able to induce the muscular phenotype under the control of the MSV LTR. Moreover, it appeared to be more difficult to obtain transgenic animals with each of the α_{sk}-actin *ski* constructions than with any of the corresponding MSV *ski* constructions.

We have measured the levels of transgene RNA levels in the α_{sk}-actin *ski* mice. As expected, the transgene is expressed in a muscle-specific manner. Although there is a considerable range in expression levels, the highest level of α_{sk}-actin *ski* mRNA is approximately 20% of the level of the MSV *ski* mRNA. The α_{sk}-actin *ski* mice do not show an overt muscular phenotype; however, when the muscles

are sectioned, some hypertrophic fibers can be seen in the animals that express the highest levels of transgene RNA. The analysis of the α_{sk}-actin *ski* mice and the additional MSV *ski* mice shows a definite correlation between the level of the transgene mRNA and the degree of the muscular phenotype.

We have also shown, in the lines of *ski* transgenic mice that develop hypertrophy, that there is a significant increase in *ski* expression in the period 10 to 12 days postpartum, which coincides with the appearance of hypertrophy. This observation raises several interesting questions. We believe, based in part on the experiments with the α-actin promoter linked to the various *ski* cDNAs, that high levels of *ski* expression are toxic early in development, which would explain the relative difficulty in producing α_{sk}-actin *ski* mice (we believe that animals that are the highest-level expressors are lost). However, the data we have obtained with the MSV/*ski* mice also show that by 10 to 12 days, high levels of *ski* expression can be tolerated, at least in muscle, and furthermore, that this high level of *ski* expression at this specific time in development can induce hypertrophy.

Human Retrovirus Section

George N. Pavlakis

Yoshifumi Adachi
Elena Afonina
Robin Bhaerman
Pantelis Constantoulakis
Emilia Hadziyannis
Jeanette Harrison
Marina Korneyeva
Markus Neumann
Mary Saltarelli
Ralf Schneider
Seong Song
James Turpin
Antonio Valentin
Agneta von Gegerfelt

The Human Retrovirus Section studies the regulation of human immunodeficiency virus (HIV) expression in human cells and the virus-cell interactions important for HIV-1 expression. Our goals include the evaluation of new methods using synergistic combinations of antiviral agents to inhibit HIV-1 replication in human cells.

A Family of Interferon-Induced Human Genes Produce RNA-Binding Proteins, Some of Which Inhibit HIV-1 Expression

We have previously reported the identification of a human interferon (IFN)-induced protein, RBP9-27, that inhibits HIV-1 expression (1992 Annual Report). We showed that RBP9-27 specifically inhibits Rev-dependent post-transcriptional steps of gene expression. This year, we continued the characterization of this interesting protein, which may define a new mechanism of IFN function against viruses.

RBP9-27 is the first cellular factor demonstrated to antagonize Rev function. The interactions of HIV-1 with the various cellular factors, including RBP9-27, have not been yet fully elucidated. These interactions may result in the restricted expression of HIV-1 in different cell types. Indeed, lentiviral replication strategy involves negative regulatory steps considered to be important for the generation of chronic active infections by these viruses. Thus, RBP9-27 probably plays a major role in the HIV-IFN interaction.

Characterization of Two Additional Members of the 1-8 Gene Family

The 9-27 gene is a member of the 1-8 gene family, which was first identified through homology to a cDNA corresponding to an IFN-inducible mRNA. At least three functional members of the 1-8 gene family are expressed in human cells: 9-27, 1-8U, and 1-8D. Using reverse transcription (RT)-polymerase chain reaction (PCR), we isolated the cDNAs of all three genes and verified their nucleotide sequence. There is 92% identity between 1-8U and 1-8D, 76% identity between 9-27 and 1-8D, and 83% identity between 1-8U and 9-27 (Figure 1). This conservation suggested that the other members of the family also encode RNA-binding proteins. To test the RNA-binding properties of the 1-8 family members, we expressed these proteins in *Escherichia coli* and performed a series of filter binding experiments using the purified proteins. As shown in Figure 2, all three proteins bind Rev-responsive element (RRE_{330}) RNA. These experiments show that all three members of the 1-8 gene family encode RNA-binding proteins that can bind RRE RNA with similar affinities. The binding affinities of these proteins for RRE were estimated to be lower than, but comparable to, that of Rev. We also demonstrated the binding properties of the 1-8 family proteins to the RRE RNA with gel retardation experiments. Using radiolabeled RRE_{330} as probe, we showed that the bacterially produced, purified proteins caused a characteristic shift indicating binding to the RNA target (Figure 3).

```
MSHTVQTFFSPVNSGQPPNYEMLKEEHEVAVLGGPHNPAPPTSTVIHIRSETSVPDHVVWSLFNTLFMNPCCL
|.|.|||.|||||||||||||||||:|||:|||||||||||||||||||||||||||||||||||||.|||
MNHIVQT.FSPVNSGQPPNYEMLKEEQEVAMLGGPHNPAPPTSTVIHIRSETSVPDHVVWSLFNTLFMNTCCL
                        |.|||:|||:||:|.....|.||||:|:|||||||||||||||||:|.|||
                      MHKEEHEVAVLGAPPSTILPRSTVINIHSETSVPDHVVWSLFNTLFLNWCCL

                                  *       *       *       *
GFIAFAYSVKSRDRKMVGDVTGAQAYASTAKCLNIWALILGILMTILLIVIPVL.....IFQAYG   1-8U
|||||||||||||||||||||||||||||||||||||||||||||:|||||::||||.....:.||..
GFIAFAYSVKSRDRKMVGDVTGAQAYASTAKCLNIWALILGIFMTILLVIIPVL.....VVQAQR   1-8D
|||||||||||||||||||||||||||||||||||||||||||:|||.::::..|:.....::|..|
GFIAFAYSVKSRDRKMVGDVTGAQAYASTAKCLNIWALILGILMTIGFILSLVFGSVTVYHIMLQIIQEKRGY 9-27

1-8U/1-8D:  96% SIMILARITY
            92% IDENTITY

1-8D/9-27:  88% SIMILARITY
            76% IDENTITY

1-8U/9-27:  89% SIMILARITY
            83% IDENTITY
```

FIGURE 1. Sequence alignment of the three 1-8 family proteins. The deduced amino acid sequences of RBP9-27, RBP1-8U, and RBP1-8D were compared using the BESTFIT program of the GCG analysis software.

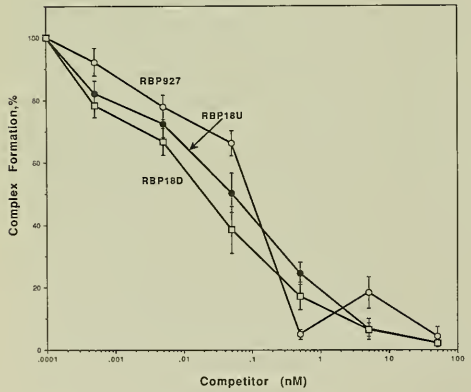


FIGURE 2. Bacterially produced, purified proteins of the 1-8 gene family were analyzed for binding to in vitro-transcribed RNA. Shown are the results from four independent filter binding experiments, in which we compared the binding characteristics of the three proteins to the RRE_{330} RNA. The mean and standard deviation from these experiments are expressed as percentages of complex formation.

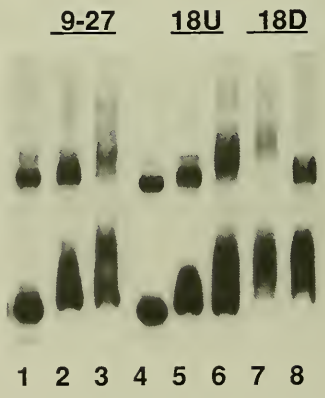


FIGURE 3. Gel retardation experiment using the bacterially produced, purified proteins of the 1-8 gene family and radioactively labeled RRE_{330} RNA. Lanes 1 and 4 show the probe alone (0.1 pmol); lanes 2 and 3 contain 25 and 40 pmol of RBP9-27, respectively; lanes 5 and 6 contain 25 and 40 pmol of RBP1-8U, respectively; lanes 7 and 8 contain 25 and 40 pmol of RBP1-8D, respectively.

We then assessed the effect of these proteins on the expression of HIV-1 in cotransfection experiments of HLtat cells. We cotransfected pL927 or pL18U plasmids [which contain the HIV-1 long terminal repeat (LTR) driving the 9-27 or the 1-8U cDNAs, respectively] with pCgagA2 and pL3rev (which contain the HIV-1 LTR driving the *gag* and *rev* genes, respectively) into HLtat cells. A dose-dependent inhibition of Gag protein expression was observed in these experiments, as is evident by Western blot analysis or by quantitation of the $p24^{gag}$ protein in an antigen-capture assay (Figure 4), while the expression of a cotransfected Rous sarcoma virus (RSV) promoter-luciferase gene was not significantly affected. However, in similar experiments involving 1-8D, cotransfection of the pL18D plasmid with pCgagA2 and pL3rev plasmids did not affect Gag protein expression (Figure 4). To further evaluate the effect of these proteins on HIV-1 expression, we cotransfected the entire molecular clone of the HIV-1 strain HXB2 with the different 1-8 family members, together with the RSV-luciferase control plasmid in HLtat cells, and measured viral protein and RNA expression. As shown in Figure 5, pL927 and pL18U specifically inhibited HIV-1 protein expression in a dose-dependent fashion, while pL18D did not affect the expression of HIV-1 at any concentration used (in three independent experiments). These results demonstrate that the three RNA-binding proteins of the 1-8 gene family affect HIV-1 expression in a different fashion.

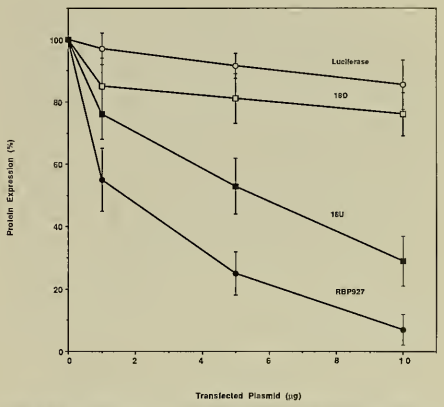


FIGURE 4. Quantitation of Gag production as measured by the $p24^{gag}$ antigen-capture assay. HLtat cells were transfected with 1 µg of the Gag-expressing plasmid (pCgagA2), 0.5 µg of the Rev-expressing plasmid (pL3crev), and increasing amounts of pL927 (solid circles), pL18U (solid squares), or pL18D (open squares). All transfections were done in the presence of the luciferase expression plasmid (pRSVluc) as control. The mean and standard deviation from three independent experiments are shown.

Effect of Coexpression of 1-8 Family Members on HIV-1

All three members of the 1-8 gene family are inducible to different levels by interferons (α, β, and γ). Because of the differential effect of the three proteins on HIV-1 expression, we performed a quantitative analysis of their expression levels in different cell types. We designed specific oligonucleotide primers that distinguish the three mRNAs and used the method of quantitative RT-PCR to assess the levels of their expression before and after interferon treatment. After normalization of the values for the level of β-actin in the sample, we compared the amount of RNA expressed by the different cell types. There was a wide variation in the basal levels as well as the IFN-induced levels of expression of the three 1-8 proteins. These findings were also confirmed by immunoprecipitations of the different

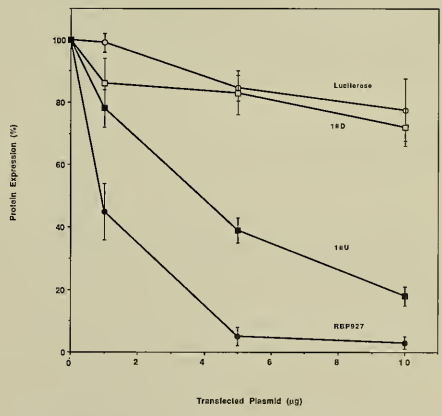


FIGURE 5. Quantitation of $p24^{gag}$ in HLtat cells transfected with 2 µg of the HIV-1 proviral clone (pHXB2) together with pRSVluc as control and increasing amounts of pL927, pL18U, and pL18D plasmids. The mean and standard deviation from three independent experiments are shown.

1-8 proteins in the same samples, using specific peptide antibodies that distinguish between the three members.

Since all three proteins of the 1-8 family coexist in different levels in various cell types, it is important to evaluate their combined effects on HIV-1 expression and propagation. To assess the effect of simultaneous expression of the different 1-8 family members on HIV-1, we performed a series of cotransfection experiments simultaneously using two members of the 1-8 gene family. As shown in Figure 6A, in three individual experiments, cotransfection of pL927 and pL18U with the pNL43 clone resulted in an additive inhibition of HIV-1 expression, as compared with transfection of either plasmid alone. Thus, transfection of 1 μg of pL927 or pL18U alone resulted in a 35% to 45% lower expression of $p24^{gag}$, whereas the same level of inhibition was achieved with 0.5 μg of each plasmid together. On the other hand, cotransfection of pL927 and pL18D in three independent experiments resulted in less inhibition, compared to pL927 alone (Figure 6B). Thus, transfection of 1 μg of pL927 resulted in a 30% to 40% reduction of $p24^{gag}$ expression, while transfection of 1 μg of pL18D did not inhibit expression; cotransfection of 1 μg of each plasmid had no significant effect on HIV-1 p24 expression, and, in fact, prevented the inhibitory effect of pL927 on HIV-1 p24 expression. The further understanding of the function of these proteins may offer new insights into virus-cell interactions and the development of novel therapeutic strategies for HIV-1 infection.

Methodology to Detect and Quantitate All mRNAs Produced by HIV-1

A quantitative RNA-PCR method able to detect all HIV-1–produced mRNAs was developed and used to study expression of different HIV-1 clones in human cells. Amplified RNAs were compared to known cDNA standards, permitting the optimization of PCR conditions and eliminating the generation of artifactual PCR bands. The results demonstrated the overall conservation of splicing among different HIV-1 clones. Although, in general, splicing is conserved, extensive qualitative and quantitative variability was observed in different clones. In addition to a universal set of RNAs, several nonuniversal mRNAs were identified, which demonstrates a great plasticity of the HIV genome, resulting in extensive splicing variability. This variability is one of the determinants of the biological characteristics of the different HIV-1 clones. The availability of a sensitive, rapid, and quantitative method to examine all HIV-1 mRNAs will allow the more detailed analysis of HIV-1 expression in human cells.

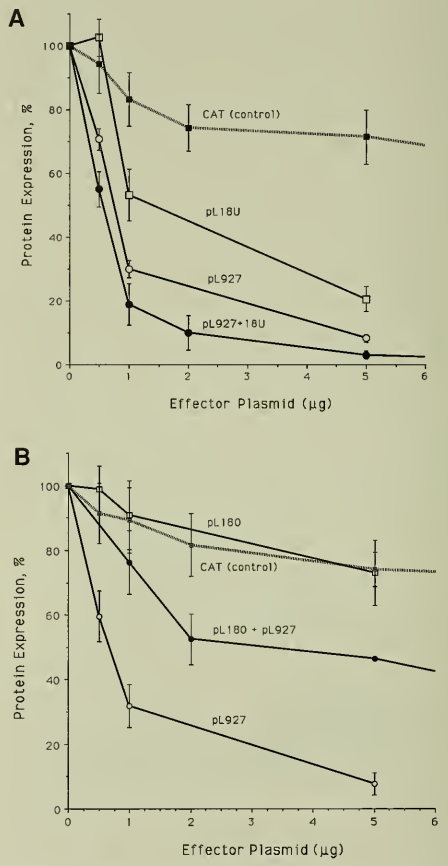


FIGURE 6. (A) Plasmids expressing RBP9-27 (pL927) and RBP18U (pL18U) were cotransfected with the pNL43 molecular clone of HIV-1, either separately (open symbols) or both at the same time (solid circles), and their effect on $p24^{gag}$ protein expression was assessed. The mean and standard deviation of three experiments are shown here, together with the CAT protein values (control, dotted line). (B) As in (A) for RBP9-27 and RBP18D proteins.

An example of the data produced by this technique is demonstrated in Figure 7, which shows our analysis of RNA from unstimulated peripheral blood mononuclear cells (PBMCs) of a patient with primary HIV-1 infection. Early after HIV-1 infection, the virus expression levels in unstimulated PBMCs are high. Within a short period of time, this expression drops at least tenfold and remains very low throughout the asymptomatic phase of infection.

Identification of Novel mRNAs Produced by HIV-1

We have previously identified additional splice sites resulting in the production of novel HIV-1 mRNAs. Some of these splice sites do not exist or are not used in all HIV-1 strains. For example, splice sites for the generation of Tev protein are not conserved in all HIVs. Since some of the additional mRNAs produce variant proteins, it is of interest to determine the role of these mRNAs on the biological properties of the various HIV-1 strains.

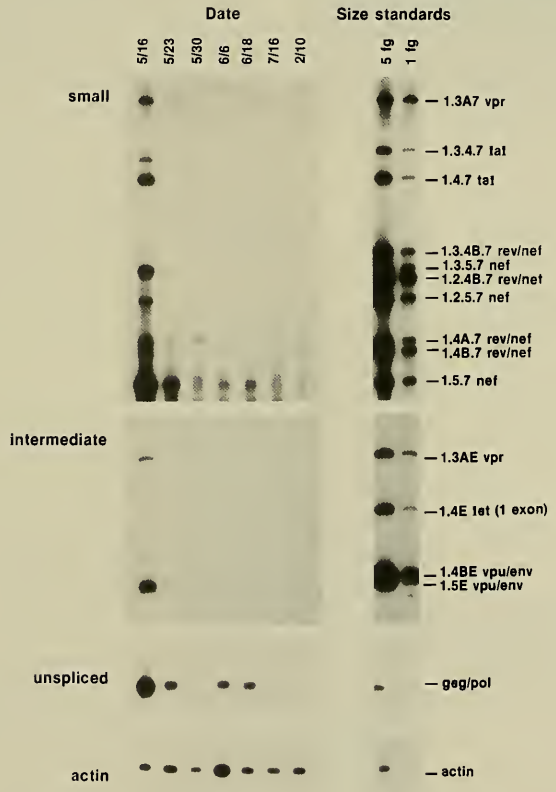


FIGURE 7. HIV-1 expression in the PBMCs of a patient with primary infection (SUMA), detected by quantitative RNA-PCR, using actin as an internal standard. After initial high levels, all the HIV-1 mRNA species drop gradually and remain very low during the asymptomatic stage of the disease.

We also used the quantitative RNA-PCR technique to identify new splice sites in HIV-1. The elimination of artifactual bands resulted in easier detection of variant mRNAs not found in all HIV-1 strains. An example of this work, both on HIV-infected cells in vitro and on cells of two different patients with primary HIV-1 infection, is shown in Figure 8. The analysis of multiply spliced species revealed several additional amplification bands in some HIV-1 strains; therefore, we decided to determine the nature of these mRNAs. One example of a novel band was detected in HIV_{LAI}-infected PBMCs (Figure 8). This band migrated below the 1.2.5.7 *nef* mRNA band and was not present in all HIV-1 strains. We cloned and sequenced this band after excision from the gel and reamplification, and showed that it is the result of a previously unidentified splicing event. This mRNA used a 3′ splice site at position 5483, upstream of the two other identified Rev splice sites. We verified that this mRNA is a Rev/Nef-producing mRNA, which we named 1.4C.7 *rev/nef*, in agreement with the terminology we have used previously. Interestingly, when we analyzed RNA directly from unstimulated PBMCs of two different patients with primary infection, we found that in the one patient (INME), the splice site 4C was used efficiently, while in another patient (SUMA), it was not (Figure 8). These results further verify the conclusion that the splicing variability (both quantitative and qualitative) found in different HIV-1 strains is extensive. In some cases, variable slicing results in the production of additional variant proteins

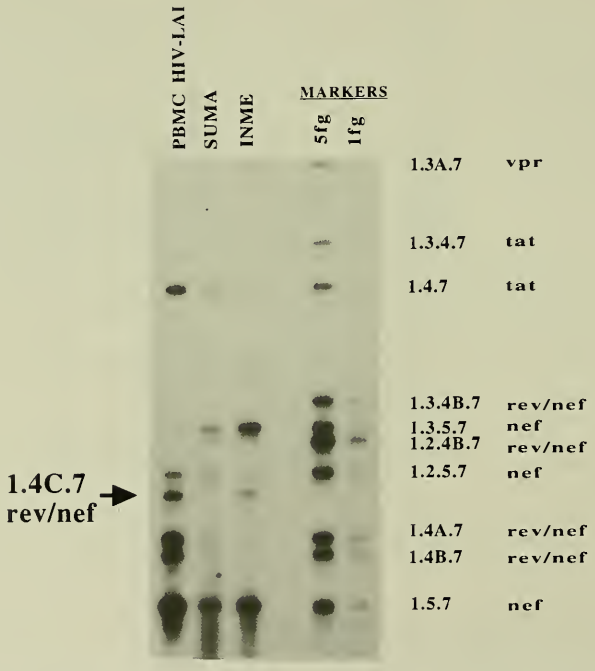


FIGURE 8. Analysis of the multiply spliced mRNAs produced in PBMCs in vitro and in unstimulated patient cells. The left lane represents amplification (30 cycles) of RNA from PBMCs infected in vitro with $HIV\text{-}I_{LAI}$, using a primer pair (BSS-SJ4.7A) that is specific for the multiply spliced species. Lines SUMA and INME are amplifications under the same conditions from unstimulated PBMCs of two patients with primary HIV-I infection. Our standard HIV-I cDNA mixture for multiply spliced species was used as markers.

such as Tev and 6DRev, while in other cases, such as in the case of 3´ splice site 4C, this splicing results in the increased production of certain viral mRNAs and proteins.

Production of p55^{gag} Precursor and Gag Particles in the Absence of Rev After Elimination of INS Elements

We have reported the elimination of negative-acting elements in the p17^{gag} region of HIV-1, which resulted in Rev-independent expression (1992 Annual Report). In continuing these studies, we have eliminated additional INS elements in the p24^{gag} and the *pol* regions of HIV-1 by introducing multiple point mutations not affecting the amino acid sequence of the encoded proteins, as previously described. This elimination allowed the expression of high levels of intact p55^{gag} precursor protein in the absence of any HIV-1 regulatory factors, including Rev. The expression was very high in human cells and led to the formation of virus-like particles (Figure 9). Interestingly, the efficient production of p55^{gag} particles in mouse cells may allow the generation of hybrid virus particles.

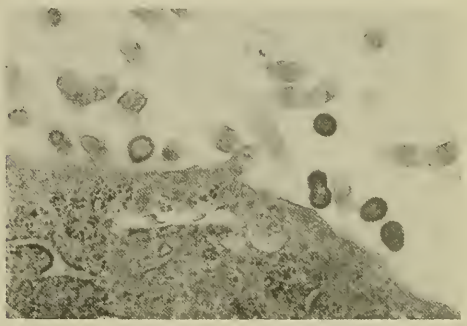


FIGURE 9. Formation of virus-like particles containing unprocessed p55^{gag} in human cells after transfection of the p55^{gag}M1-10 expression vector in the absence of Rev protein.

Virus Production in the Absence of Tat in Several Cell Lines

The transcriptional activator Tat is an important regulatory HIV protein. Although Tat is important, it has been shown that several cellular factors are able to activate expression of the HIV LTR in the absence of Tat. To examine whether different cell lines can support *tat*-defective HIV replication, we generated *tat*-defective viral stocks by using site-directed mutagenesis. We produced two *tat*-defective stocks (HIVtat3, HIVtat10) by the introduction of either three or ten point mutations, respectively, into the region of the *tat* gene. HIVtat3 mutations created one stop codon in the *tat* region, while HIVtat10 created seven stop codons. The infectivity of *tat*-defective virions produced from transfected HLtat cells was assayed in different cell lines. Persistent viral expression was detected in T-lymphoblastoid cell lines ZS-1 and HPBALL, and the monocytoid cell line THP-1 (Figure 10). We observed that *tat*-minus virus production was about two to three orders of magnitude less than that of wild-type virus in the same cell lines. A detectable level of virus production was maintained for up to 8 weeks. A low or undetectable level

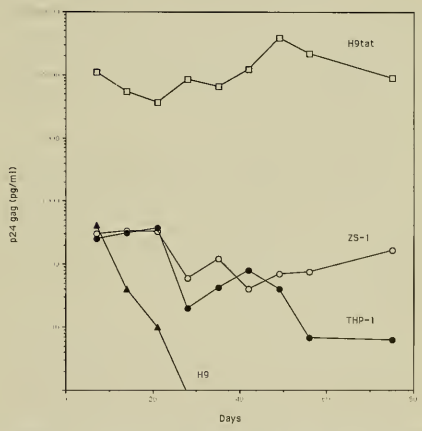


FIGURE 10. Virions produced from HLtat cells transfected with the *tat*-defective strain HIVtat10 were used to infect different cell lines. The virus expression by these cell lines was measured by the p24^{gag} antigen-capture assay. ZS-1, T-lymphoblastoid cells (open circles); THP-1, monocytoid cells (solid circles); H9, T-cells (solid triangles); H9 tat, T-cells producing Tat proteins (open squares).

of *tat*-defective virus expression was found in H9, U937, Raji, and Molt4 cell lines. In these cell lines, *tat*-defective virus production was gradually reduced to undetectable levels 3 to 4 weeks after infection, and the viral genome could no longer be detected by PCR.

To exclude the possibility that *tat*-defective virus was reverted to *tat*-positive virus during propagation in the cell lines, we used PCR and cloning to analyze the DNA sequence of the *tat* gene from cell lines infected with *tat*-defective virus. After sequencing 25 different clones of HIVtat10, we failed to detect any revertants to the *tat*-positive genotype. This result is in contrast to experiments with HIVtat3, in which revertants to the *tat*-positive genotype were shown both by sequencing and by functional assays after propagation in different cell lines.

In addition, we assayed the viral stocks for Tat activity in appropriate indicator cell lines. Passage of this virus to indicator cell lines developed in our laboratory, such as U38, demonstrated that no transactivation of the LTR occurred. This observation suggests that cellular factors in certain cells are responsible for partially replacing Tat function, which, taken together with other data, further suggests that monocytes are able to direct expression of the LTR promoter in the absence of Tat. The combination of factors that allows virus propagation in the absence of Tat has not yet been identified.

In some cell lines, *tat*-minus virus expression can be further activated by different cytokines and phorbol myristate acetate (PMA) (Figure 11). Costimulation by PMA and interleukin-6 (IL-6) or granulocyte-macrophage colony-stimulating factor (GM-CSF) resulted in an increased expression of *tat*-defective virus in ZS-1 and THP-1 cells, but not in HPBALL cells.

Our studies of *tat*-defective proviruses suggest that *tat*-defective genomes can be maintained in certain cell types for long periods of time, during which they continue to produce low titers of infectious virus. Furthermore, *tat*-defective virus may be activated by different stimuli leading to virus production. The defective genomes may revert to *tat*-plus, resulting in more efficient viral propagation. This possibility suggests a role for *tat*-defective genomes during infection.

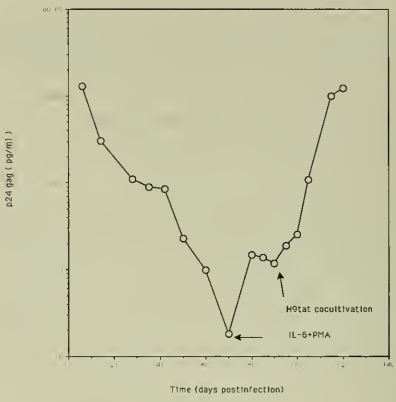


FIGURE 11. Induction of *tat*-defective provirus (HIVtat10) expression by PMA and IL-6 treatment. At 70 days postinfection, PHT1 cells infected by HIVtat10 were induced by PMA and IL-6. This treatment resulted in a tenfold activation of expression. *Tat*-defective virus was readily transmitted to H9 tat cells by cocultivation, resulting in high levels of virus expression.

Identification of Rev and RRE of Caprine Arthritis-Encephalitis Virus (CAEV)

Lentiviruses are a subfamily of nononcogenic retroviruses that cause chronic progressive diseases. These lentiviruses target cells of the monocyte/macrophage lineage. CAEV is a lentivirus that induces arthritis and mastitis in adult goats and leukoencephalomyelitis in young animals. CAEV is closely related to visna virus, the prototype nonprimate lentivirus.

The genomes of all lentiviruses contain genes that regulate the lentivirus gene expression; one of these is *rev*, a gene that regulates the expression of viral proteins via posttranscriptional mechanisms. The ovine and caprine lentiviruses have a common genomic organization, which is simpler than that of HIV-1. CAEV and visna virus are predicted to encode three nonstructural proteins, whereas HIV-1 has been shown to produce seven. A cDNA clone was isolated from CAEV-infected cells and shown to encode the 18-kilodalton Rev protein of CAEV. Experiments using antibodies against CAEV Rev (Rev-C) demonstrated that the CAEV Rev protein accumulated in the nucleus and, in particular, in the nucleolus of transiently transfected cells. A 300-base pair fragment (nucleotides 7850 to 8150) spanning the region encoding the N terminus of the envelope transmembrane protein and containing a highly structured RNA element was substituted for the HIV-1 RRE in an HIV-1 Gag expression vector. Expression of the Gag protein was dramatically increased when Rev-C was added in trans, indicating that this fragment contained

the cis-acting CAEV RRE (RRE-C). Cross-activation by the Rev/Rex proteins of other lentiviruses and members of the HTLV-I family indicated that this RRE could interact with Rev or Rex proteins of other viruses. This result suggests that highly divergent lentiviruses share similar mechanisms and cofactors regulating posttranscriptional viral gene expression. Thus, the Rev/RRE mechanism appears to be the most conserved regulatory mechanism in lentiviruses and other complex retroviruses.

Human Retrovirus Pathogenesis Group

Barbara K. Felber

Patricia Carney
Christian Cunningham
Georgios Nasioulas
Carlos Tabernero
Andrei S. Zolotukhin

Our goal is to study the cellular mechanisms involved in the posttranscriptional regulation of viral and cellular mRNAs. The human immunodeficiency virus type 1 (HIV-1) and the human T-cell leukemia virus type I (HTLV-I) use elaborate posttranscriptional regulation to express the subset of viral mRNAs encoding the structural proteins. This regulation involves a combination of both positively and negatively acting viral and cellular factors. We are also studying the cross-rescue of posttranscriptionally regulated mRNAs by viral and cellular mechanisms. To establish a model system, we compared the positive regulatory system of the transferrin receptor (TfR) mRNA with those of HIV-1 and HTLV viruses. Since the regulatory factors Rev and Rex can overcome posttranscriptional defects of viral or cellular origins at multiple steps (stability, transport, and translation of mRNAs), they may represent rescue systems with a more general application.

RBP927 is a member of the interferon-inducible gene family 1-8 and has been shown to inhibit HIV-1 expression. We have been interested in the antiviral activity of RBP927 and other members of the 1-8 gene family against other lentiviruses and the HTLV family of retroviruses. Efficient binding of these factors to the corresponding RNA targets of Rev and Rex has been shown to be a necessary requirement for interference with Rev/Rex function. We are interested in the antiviral activity against other viruses and in the identification of the cellular function of these factors.

The identification of RBP927 as an inhibitor of Rev-mediated gene expression of HIV-1 led us to design new recombinant retroviral vectors for gene transfer experiments into CD4-expressing cell lines. The presence of RBP927 severely affects virus propagation. We are applying these findings in the development of novel approaches towards gene therapy for AIDS.

Identification of Inhibitory/Instability (INS) Elements Within the gp120 and gp41 Genes

We previously identified regions within *env* mRNA that are responsible for inefficient expression in the absence of Rev. To obtain a more detailed understanding of the elements involved, we designed the following approach. The *env* coding region was subdivided into different consecutive fragments. These fragments and combinations thereof were amplified by polymerase chain reaction (PCR) using synthetic oligonucleotides and inserted downstream of the mutated HIV-1 $p37^{gag}$ gene in the vector p37M1-10D. p37M1-10D expresses the $p37^{gag}$ gene in a Rev-independent manner, since all of its inhibitory elements were eliminated. One day after transfection of HLtat cells with these constructs, the cells were harvested and analyzed for *gag* expression. Figure 1 shows that the presence of the individual fragments (1 to 6) as well as the combination of fragments 1+2 marginally affected *gag* expression. In contrast, the combination of fragments 1+2+3 and 4+5e drastically lowered *gag* expression. These results suggested that the INS elements are spanning a region of fragments 2/3 and 4/5e. We designed synthetic oligonucleotides to mutagenize these regions. Experiments are in progress to introduce point mutations that alter the nucleotide composition of these regions. Our long-term goals are to determine the

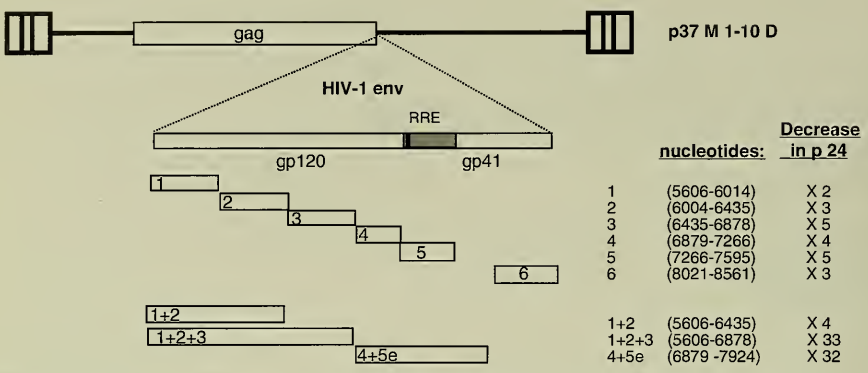


FIGURE 1. The identification of INS elements within gp120 and gp41, using the p37M1-10D test system. Schematic structure of the *env* open reading frame. Different fragments (1 to 6) of *env* were PCR amplified and inserted into the unique restriction sites in the polylinker located downstream of the p37M1-10 *gag* gene in p37M1-10D. HLtat cells were transfected with these constructs. One day later, the cells were harvested and $p24^{gag}$ production was determined by antigen-capture assay. The expression of Gag from these plasmids was compared to Gag production of p37M1-10D. Fragments 5 and 4+5e contain the RRE and respond to Rev.

underlying cellular mechanism responsible for this defect and to eliminate the downregulatory effect of these elements. We are also interested in designing novel retroviral vectors. We believe that the use of HIV-1 *env* may improve targeting of recombinant retrovirus particles to CD4+ human cells. The ability to express HIV-1 *env* in a Rev-independent manner is necessary to achieve this goal.

Cross-Rescue of Posttranscriptionally Regulated mRNAs by Viral and Cellular Mechanisms

The expression of the TfR mRNA and the structural proteins from HIV-1 and HTLV-I is controlled by post-transcriptional regulatory mechanisms involving both positive and negative elements. In these viral and cellular systems, defined RNA elements responsible for mRNA defects are the INS elements (i.e., INS-1) and the instability determinant within the iron-response region (IRR), respectively. In both systems, the regulatory factors [Rev, Rex, or iron-responsive element binding protein (IRE-BP)] interact directly with an RNA element [Rev-responsive element (RRE), Rex-responsive element (RXRE), or IREs], resulting in stabilization and efficient expression of the corresponding mRNAs. To investigate whether this strategy implies common pathways of mRNA utilization, we have studied expression from hybrid mRNAs that contained the viral or the cellular instability determinant and the binding sites of the regulatory proteins Rev, Rex, and/or IRE-BP. Last year, we reported that the cellular stabilization system for TfR mRNA can act on unstable HIV-I mRNAs and, conversely, that the Rev/RRE interactions can counteract the cellular mRNA instability element within the TfR mRNA.

The IRE-BP had been shown to stabilize the TfR mRNA via direct binding to IRE elements within IRR, which is thought to prevent association of putative cellular destabilizing factors with the instability determinant located within the same element. The hybrid mRNAs studied contain the viral INS-I element as an additional strong instability determinant located about 400 nucleotides 5′ of the IRR. Since these hybrid mRNAs accumulate in the presence of active IRE-BP to the same extent as the TfR mRNA, we conclude that the TfR stabilization mechanism is not limited to the protection of the IRR instability determinant from putative IRR-specific destabilization factors. The mechanism that permits the rescue of the instability element distant from the site of IRE-BP/IRE interaction remains to be elucidated. The IRE-BP/IRE interaction only partly relieved the inhibitory effect of INS-1, as the increase in the mRNA accumulation was not

followed by the increase in $p17^{gag}$ synthesis. We speculate that although IRE-BP exhibits Rev-like mRNA stabilization activity, it cannot substitute for Rev to counteract the translational defect caused by INS-1 in RD4 cells.

We have also demonstrated that Rev can rescue the expression of hybrid mRNAs that contain the IRR as the only strong instability determinant and the viral positive RNA elements (RRE). We expanded our studies to other posttranscriptional viral rescue systems such as the RXRE of HTLV-I and its positive regulator, the Rex protein (Figure 2). We inserted the RXRE into plasmid p17M, which expresses the INS-depleted version of the $p17^{gag}$ gene in a Rev-independent manner. HLtat cells were transfected with the resulting plasmid (p17MX) in the absence or presence of a Rex-expressing plasmid. Although the presence of the RXRE lowered *gag* expression slightly (threefold) when compared to p17M, the Rex protein did not significantly affect expression from p17MX (1.5-fold increase). We then inserted the IRR as the only instability determinant in p17MX, resulting in p17MX.I. The presence of the IRR lowered *gag* expression from the hybrid mRNA sixfold, while the presence of Rex increased *gag* expression eightfold, i.e., to the level of p17MX. These results demonstrate that the effect of the cellular instability element IRR could also be counteracted by the interaction of the regulatory Rex protein with the RXRE. Since cellular factors are responsible for the downregulation of the viral mRNAs, these results suggest that the study of Rev and Rex function will aid the understanding of the regulation of cellular mRNAs.

Our studies provided the first demonstration that viral mRNA rescue mechanisms can counteract a defect caused by a cellular instability element. Therefore, these data suggest that the underlying mechanisms causing the defect may be multiple and interchangeable (Figure 3). The Rev and Rex factors of the different HIV/HTLV viruses had been shown to be interchangeable in many, but not all, cases. The interchangeability of these factors requires their efficient binding to the heterologous RNA target elements. As a result, the various Rev/Rex factors also counteract the effects of the heterologous viral INS elements. We had identified several INS elements in HIV-1, HTLV-I, and simian immunodeficiency virus (SIV). These elements are distinct in nucleotide sequence and can act as INS elements in a heterologous mRNA. Since Rev and Rex can overcome heterologous viral and cellular (as shown here with the example of TfR) posttranscriptional defects, they may represent rescue systems with general application. We are conducting experiments to test this hypothesis. This model is further corroborated by the fact that the

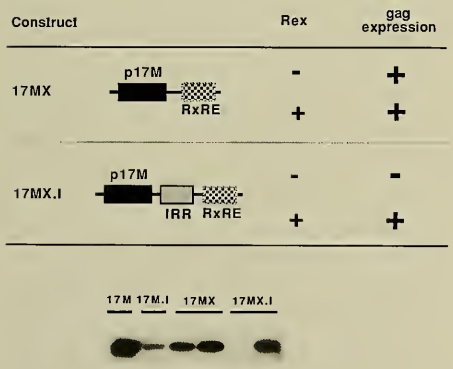


FIGURE 2. The Rex-RXRE interaction counteracts the negative effect of the IRR. HLtat cells were transfected with the indicated constructs shown in the top panel in the absence or presence of a Rex-expressing plasmid. The next day, the cells were harvested and total protein was extracted. The $p17^{gag}$ protein was visualized on a Western blot, using serum from an HIV-1–infected human patient.

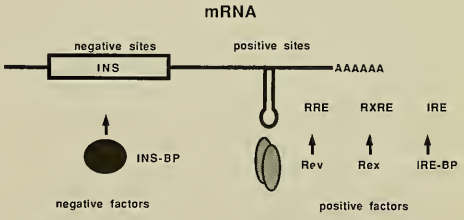


FIGURE 3. Cellular or viral rescue mechanisms can counteract multiple negative RNA determinants. Cellular factors (INS-BP) are thought to bind to negative-acting elements (INS) on the mRNA, which results in posttranscriptional downregulation. Positive-acting factors of viral (Rev, Rex) or cellular (IRE-BP) origin can counteract the negative determinants. These factors act by binding to distinct sites on the RNA (RRE, RXRE, and IRE).

Rev/Rex factors are associated with several posttranscriptional steps. Rev, the best-studied factor, has been shown to affect the half-life, transport, and translation of RRE-containing viral mRNAs. In this model, Rev is thought to chaperone RRE-containing mRNAs through the entire pathway from the nucleus to the cytoplasm.

RBP927 Inhibits Rex-Mediated mRNA Expression of HTLV-I

The interferon α, β, and γ-inducible gene RBP927 binds to the RRE of HIV-1 and inhibits Rev-mediated gene expression. The RRE is proposed to form a complex secondary structure consisting of a stem and five hairpins. Rev interacts with the left-hand portion of the structure (i.e., the region encompassing hairpin 1). Our in vitro binding experiments revealed that RBP927 binds to the right-hand portion (the region encompassing hairpins 3, 4, and 5). Coincidentally, this is the same region of the RRE that contains the binding site for Rex of HTLV-I, as we had shown previously. We also demonstrated the interchangeable function of Rev and Rex on the RRE. We designed experiments to study the effect of RBP927 on Rex-mediated gene expression from RRE-containing experiments in vivo. Cotransfection experiments in HLtat cells confirmed that RBP927 effectively inhibits Rex function.

This finding led us to investigate whether RBP927 can bind to RXRE RNA and thereby interfere with Rex function. In vitro binding/competition experiments revealed that RBP927 binds to RXRE RNA with an affinity similar to that of RRE. We then tested the function of RBP927 in cotransfection experiments in human cells. The intact molecular clone of HTLV-1 was cotransfected with increasing amounts of the expression vectors for RBP927. One day later, the cells were harvested and analyzed for *gag* expression (Figure 4). RBP927 inhibited the expression of HTLV-I in a dose-dependent manner. Additional experiments verified that RBP927 interferes with the mechanisms of function of Rex. These results are similar to the study on HIV-1 conducted by the Human Retrovirus Section. Therefore, these factors act similarly on different RNA targets of distantly related retroviruses by binding to the RNA target of the regulatory Rev/Rex factors, which results in interference in the function of the regulatory factors. The underlying mechanisms are currently under investigation.

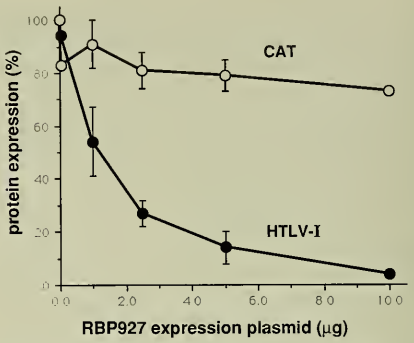


FIGURE 4. RBP927 inhibits HTLV-I expression. HLtat cells were transfected with 2 μg of the intact molecular clone of HTLV-I (pCS-1) in the presence of increasing amounts of the RBP927-expressing plasmid pL927. The addition of the chloramphenicol acetyltranferase (CAT)-expressing plasmid (pCMVCAT) serves as a control for transfection efficiency and acts as an internal indicator for nonspecific inhibition. One day later, Gag expression was measured using the $p24^{gag}$ antigen-capture assay. CAT expression was measured using the CAT enzyme-linked immunosorbent assay (ELISA). The Gag production in the absence of pL927 is considered 100%.

RBP927 and RBP18U Interfere with Rev/Rex Function in Different Complex Retroviruses

RBP927 interferes with the function of both Rev of HIV-1 and Rex of HTLV-I by binding to the corresponding RNA targets RRE and RXRE, respectively. The family of interferon-inducible 1-8 genes to which RBP927 belongs consists of several genes and pseudogenes. Studies performed by the Human Retrovirus Section revealed that two other members (RBP18U and RBP18D) bind to the RRE of HIV-1 in vitro. Functional analysis of these factors revealed that while RBP18U also interferes with Rev function, similar to RBP927, RBP18D does not affect HIV-1 expression.

In a collaborative effort with the Human Retrovirus Section, we investigated whether these factors (RBP927, RBP18U, and RBP18D) interfere with the mechanisms of Rev/Rex function of other lentiviruses such as HIV type 2 (HIV-2), SIV, and caprine arthritis encephalitis virus (CAEV), as well as the HTLV family of retroviruses such as HTLV-1 and bovine leukemia virus (BLV). In vitro binding/competition experiments revealed that all these factors

bind to the tested RRE/RXRE elements, albeit with different affinities. In vivo studies were performed by cotransfecting intact molecular clones of HIV-2, SIV, HTLV-I, and BLV with increasing amounts of the expression vectors for RBP927, RBP18U, and RBP18D. The study on CAEV was performed with a subgenomic hybrid plasmid that contained the HIV-1 *gag* gene linked to the CAEV RRE in the presence of the CAEV Rev expression plasmid. The data of this analysis are summarized in Table 1. Both RBP927 and RBP18U inhibit the expression of CAEV, HTLV-I, and BLV in addition to that of HIV-1, but they do not affect HIV-2 and SIV expression. RBP18D seems to be inactive on all the viruses studied so far. Additional experiments confirmed that inhibition by these factors affects only the function of the Rev and Rex factors. A comparison of the in vitro binding/competition study and the in vivo functional analysis revealed a tight correlation between binding affinities and function. Therefore, our data suggest that efficient binding of these factors is necessary for function. All these experiments were performed in the same cellular background. Therefore, we can also rule out that these factors affect an important cellular component necessary for Rev/Rex-mediated expression, since HIV-2 and SIV expression were not affected. It is interesting to note that RBP927 and RBP18U mediate their function by interfering specifically with the mechanisms of function of the Rev/Rex factors. The underlying mechanisms leading to inactivation of Rev/Rex function are currently under investigation.

TABLE 1. Effects of RBP expression

Virus	RBP927	RBP18U	RBP18D
HIV-1	+	+	–
HIV-2	–	–	–
SIV	–	–	–
CAEV	+	+	–
HTLV-I	+	+	–
BLV	+	+	–

Gene Therapy Approaches Using Novel Recombinant Retroviral Vectors for the Inhibition of HIV-1 Expression

Studies on the Rev-mediated expression of HIV-1 have revealed two antagonists with potential practical applications. Previously, we had reported that certain Rev mutants displayed a transdominant phenotype. Recently, the Human Retrovirus Section identified RBP927 as the first cellular antagonist of Rev function. We are applying these findings to the development of novel approaches towards gene therapy for AIDS.

We have generated recombinant retroviral vectors expressing the potential Rev-antagonists from two types of internal promoters: the inducible HIV-1 long terminal repeat (LTR) promoter and the constitutive-expressing human cytomegalovirus (HCMV) promoter. These recombinant viruses were used to infect CD4-expressing cell lines from which G418-resistant mass cultures were generated. To study the antiviral potency of the transduced genes, we subjected the cells to infection by HIV-1 and determined virus production by measuring $p24^{gag}$ levels. We had previously inserted the transdominant *rev* gene into such recombinant retroviral vectors. Together with Seong Song and Marina Korneyeva of the Human Retrovirus Section, we have generated recombinant retroviruses expressing RBP927. Upon challenge with HIV-1, virus production is inhibited by at least 90% (Figure 5).

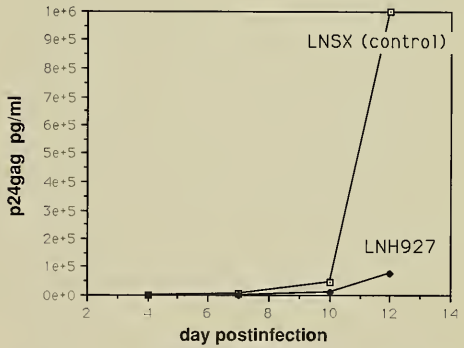


FIGURE 5. Inhibition of HIV-1 propagation in cells expressing RBP927. A recombinant retrovirus containing the RBP927 gene linked to the HIV-1 LTR promoter was used to infect a pre-T cell leukemic cell line (HPB-ALL). A mass cell culture harboring the RBP927 construct was infected by HIV-1/LAI. Cells containing the parental recombinant retrovirus served as a control. Culture supernatants were analyzed at different time points after infection for virus production, using the $p24^{gag}$ antigen-capture assay.

However, this protection appears to be only partial, since virus propagation can be observed after a longer incubation.

Since complete protection by a single agent is unlikely to be achieved, we wish to study combinations of inhibitory factors in an effort to potentiate the effects of the different factors in inhibiting HIV-1 propagation. We are currently exploring the effectiveness of combining transdominant Rev and RBP927. In addition, we are testing the combinational effect of other inhibitory factors acting at different steps in the virus life cycle. Examples of such candidates are the transdominant mutants of Gag. Their presence has been reported to render particles noninfectious. In order to combine the transdominant Gag approach with our effort to interfere with Rev function, *gag* must be expressed in a Rev-independent manner. For this purpose, we took advantage of the INS-depleted version of the *gag* gene generated by the Human Retrovirus Section. We introduced several mutations into the region of $p17^{gag}/p24^{gag}$ and are currently testing these mutations for their inhibitory effect. This approach gives us the unique opportunity to study the antiviral effect of these mutants in combination with transdominant Rev and RBP927.

To achieve efficient expression of several genes in recombinant retroviral vectors, we are currently testing several approaches. In the first approach, the introduced genes are expressed from different independent promoters. In the second approach, bicistronic expression vectors have been generated. To this end, we took advantage of our previous extensive study on the determinants that allow bicistronic expression, using the examples of the HIV mRNAs. We had demonstrated that the *rev* mRNA expresses both Rev and Nef and that the *vpu/env* mRNA expresses Vpu and Env. The underlying mechanism that allows expression of two proteins from one mRNA is leaky scanning. In contrast, the *tat* mRNAs are monocistronic. We demonstrated that the major determinant rendering an mRNA mono- or bicistronic is several nucleotides upstream of the translation initiation codon AUG. Based on these studies, we designed bicistronic mRNAs that are optimized for efficient expression of both genes. Transient expression experiments into human cells verified the efficient expression of multiple genes. We have generated cell lines with the different recombinant retroviruses and verified proper transfer of the transgenes. We are currently investigating the antiviral effect of these factors alone and in combination with HIV.

Eukaryotic Transcriptional Regulation Group

Peter F. Johnson

Mark E. Baer
Carrie A. Cantwell
A. Jeannine Lincoln
Yury Monczak
Ashmi Patel
Esta Sterneck
Tung Tran
Deborah Wancio
Simon C. Williams

The central problem of developmental biology is to understand how the diverse cell types and tissues of an adult organism arise from a single progenitor cell. It is generally accepted that during early development, a series of programmed and inductive events cause a multipotential cell population to diverge into a number of independent lineages. The establishment of cell lineages precedes the terminal differentiation step at which definitive, specialized cell types are produced. Cell differentiation, which usually occurs in response to specific extracellular signals, culminates in the expression of a set of specialized genes in the mature cell. The transcription of these cell-specific genes depends on the synthesis or activation of a new set of regulatory proteins in the nucleus. The study of cell type-restricted transcription factors thus provides a useful starting point from which to investigate the genetic control of cell lineage commitment and terminal differentiation.

The Eukaryotic Transcriptional Regulation Group has a long-term goal of understanding (1) the genetic switches that specify the maturation of committed precursor cells into terminally differentiated cells, as well as the controls that underlie prior cell lineage decisions, and (2) the mechanisms by which specialized genes are activated in differentiated cell types. As a means of addressing these and related problems, we have been investigating a family of DNA-binding proteins that appear to function in the establishment and/or maintenance of differentiated cell phenotypes. These proteins are related in amino acid sequence and DNA-binding specificity to the tissue-specific regulatory protein C/EBP (*C*CAAT/*e*nhancer *b*inding *p*rotein), a member of the basic region:leucine zipper (bZIP) class of DNA-binding proteins (Figure 1). Each of the C/EBP proteins is expressed in a unique set of differentiated cell types, suggesting that these proteins are involved in regulating differentiation or lineage commitment of certain specialized cells.

Our studies of the C/EBP-related *p*roteins (CRP1, CRP2, CRP3, and C/EBP) involve both biochemically and biologically oriented projects that can be divided into the following major areas: (1) the roles of CRPs in controlling differentiation and cell-specific gene expression in hepatic, hematopoietic, and neuronal cell systems; (2) the properties that endow each CRP with a unique regulatory function, despite the identical DNA-binding specificities of all family members; (3) the basis for the developmental and tissue-specific regulation of *crp* gene expression; (4) the potential coupling of CRP expression with growth arrest in terminally differentiating cells; (5) cell-type specificity of transcriptional activation by C/EBP and CRP2; and (6) the domain organization and structures of the CRPs. Our progress on two of these projects will be reviewed in this report.

Cell-Specific Transactivation by C/EBP and CRP2

CRP2 and C/EBP Activate Transcription Preferentially in Hepatic Cells. C/EBP and CRP2 (also described as NF-IL6, LAP, IL-6DBP, AGP/EBP, and C/EBPβ) are expressed at high levels in adult liver, and a number of liver-specific genes, including albumin, contain C/EBP binding sites in their promoter or enhancer regions. C/EBP and CRP2 are

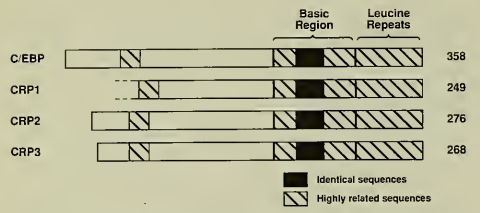


FIGURE 1. Amino acid sequence relatedness among C/EBP family members. All members share a highly related bZIP region and a second, less homologous domain near the N terminus.

therefore thought to regulate the transcription of many liver-specific genes. In previous experiments, C/EBP overexpression was found to enhance transcription from the albumin promoter at least 50-fold in a human hepatoma cell line (HepG2) but caused only a modest stimulation in L fibroblasts (Friedman et al., Genes Dev 3:1314, 1989). We subsequently compared the abilities of CRP2 and C/EBP to stimulate transcription of a cotransfected target gene, pAlbCAT, which contains a reporter gene (chloramphenicol acetyltransferase, CAT) fused to the mouse albumin promoter (1991 Annual Report). When C/EBP or CRP2 expression plasmids were transfected into HepG2 and HeLa cells together with pAlbCAT, transcription from the reporter gene increased 35- to 40-fold in HepG2 cells, whereas in HeLa cells the maximal enhancement of pAlbCAT by C/EBP was only 6.5-fold and 2.5-fold by CRP2. This effect of the recipient cell type was not due to reduced synthesis of the activator proteins in HeLa cells, since in this experiment the expression of C/EBP and CRP2 in HeLa cells was nearly fivefold greater than in HepG2 cells. These experiments demonstrate that both C/EBP and CRP2 are approximately ten times more potent as transactivators in HepG2 cells than in HeLa or L cells. The acidic activator protein GAL4-VP16 is highly active in all cell lines (Figure 2), showing that cell-type specificity is restricted to certain regulatory proteins.

Identification of a Cell-Specific Regulatory Domain Within CRP2. We observed cell-specific transactivation by C/EBP and CRP2 when an artificial promoter composed only of the albumin C/EBP binding site and core promoter was used to drive the CAT gene (Figure 2). This result demonstrates that the differential activity of these factors is not due to interactions with cell-specific proteins bound at neighboring sites within the albumin promoter. Rather, the differences probably result from the cell-specific modulation of functions intrinsic to CRP2, such as DNA binding or transcriptional activation. We sought to identify the domains within C/EBP and CRP2 that are responsible for cell specificity. By constructing domain switch proteins with the yeast bZIP protein GCN4, we determined that the sequences responsible for cell specificity in C/EBP and CRP2 are external to the DNA-binding domain and reside in the N-terminal half of the protein. The N-terminal regions of C/EBP and CRP2 were previously shown to contain the transcriptional activating sequences in these proteins. Deletion mutants in CRP2 identified an activation domain near the N terminus of the protein that corresponds to a region of homology among all members of the C/EBP family (see Figures 1 and 3). We have fused this element from each of the CRPs to the GAL4 DNA-binding domain and have shown that the resultant chimeras activate transcription from a GAL4-dependent promoter, thus demonstrating that the conserved region functions as an activation domain. These chimeras activate transcription equally well in HepG2 and L cells, indicating that the CRP

		Recipient Cell Line		
		HepG2	L	HeLa
Activator Protein	C/EBP	26.1 ± 6.1	3.6 ± 2.3	6.3 ± 3.2
	CRP2	14.0 ± 3.9	1.8 ± 0.7	2.7 ± 2.5
	GAL4-VP16	>53	38	>50

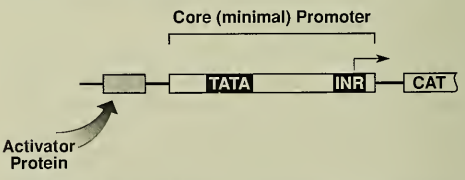


FIGURE 2. Transactivation by C/EBP and CRP2 is cell type-dependent. Synthetic promoters containing a minimal promoter element fused to either C/EBP or GAL4 binding sites were inserted upstream of the CAT gene (bottom panel). The constructs were transfected into the indicated cell line in the presence or absence of expression vectors for C/EBP, CRP2, or GAL4-VP16. The fold activation resulting from expression of the appropriate activator protein was determined in duplicate experiments, and the mean values and standard deviations are presented.

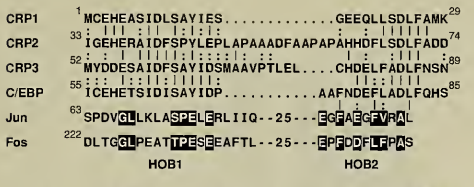


FIGURE 3. The activation domain of CRP2 corresponds to a bipartite region of homology shared by all of the C/EBP-related proteins. The C-terminal half of the homology region is also related to the HOB2 element of a bipartite activation domain in Jun and Fos. The relationship between HOB2 and the corresponding sequence in C/EBP was described by Konzarides and colleagues (Genes Dev 6:1810, 1992), who also showed that the C/EBP "HOB2" element could functionally replace Jun HOB2. Amino acid numbers for each protein segment are indicated.

activation domains are not intrinsically cell specific. Cell specificity thus requires that the activation domain reside in its natural context in the protein.

Interestingly, certain internal deletion mutants in CRP2 alleviate cell specificity. The removal of sequences immediately N-terminal to the DNA-binding domain increases CRP2 activity in L cells to a level nearly equal to that in HepG2 cells (Figure 4). From this result, we infer that a negatively acting regulatory domain resides in the region adjacent to the DNA-binding domain that inhibits CRP2 activity in nonpermissive cells. This inhibition could involve either the DNA-binding or transactivation functions of

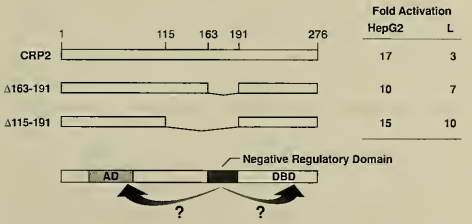


FIGURE 4. The cell-specific activity of CRP2 is controlled by a negative regulatory domain. Two internal deletions (Δ163-191 and Δ115-191) substantially alleviate the difference in transactivation between HepG2 and L cells. The arrows signify potential interactions between the regulatory region and other domains in CRP2. AD, activation domain; DBD, DNA-binding domain.

the protein, or both. We have obtained preliminary data from bacterially expressed proteins indicating that deletion of the regulatory domain may stimulate the DNA-binding activity of CRP2. This result supports previous reports that N-terminal truncations of CRP2 enhance its DNA-binding activity (Poli et al., Cell 63:643, 1990; Descombes and Schibler, Cell 67:569, 1991).

To test whether the activating region is also affected by the regulatory domain, we constructed chimeric activator proteins containing the GAL4 DNA-binding domain fused to the N terminus of a series of CRP2 mutants deleted progressively from the C terminus. We examined these proteins for transactivation of a GAL4-dependent promoter in both cell types. The results (Figure 5) indicate that both the basic region and the regulatory domain have negative effects on transactivation (the inhibition contributed by the leucine zipper appears to be a general effect of having a second dimerization domain in the GAL4 chimeras; data not shown). Because these proteins are tethered to the target promoter via the heterologous GAL4 DNA-binding domain, we conclude that the CRP2 regulatory domain, and to a lesser extent the basic region, must function to inhibit the CRP2 activation domain. However, each of the GAL4-CRP2 chimeras activates to a similar extent in HepG2 and L cells, regardless of whether it is a strong or weak activator. This result suggests that cell specificity in some way requires binding to DNA via the bZIP DNA-binding domain (see model below).

Taken together, our results suggest that CRP2 adopts a highly folded structure in which both the activation and DNA-binding domains are inhibited by interactions involving the central regulatory region. A speculative model for the structure of CRP2 is depicted in Figure 6. We presume that the "condensed" form of CRP2 represents the default state of the protein and that in permissive cells such as HepG2, either a protein modification or the binding of another protein activates the DNA-binding and transactivation functions. The protein becomes activated due to the disruption of folding interactions between the regulatory region and the other two domains. However, this model fails to explain the lack of cell specificity of GAL4-CRP2 hybrid proteins that contain the regulatory domain, since these hybrids should be fully active in HepG2 cells but not in L cells. Therefore, we propose that CRP2 undergoes a second activation step upon binding to DNA via the

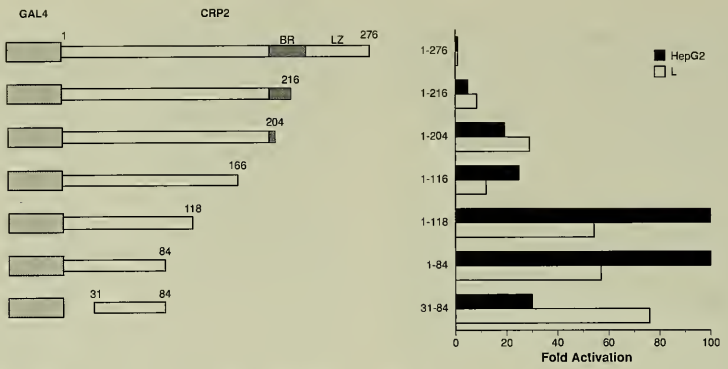


FIGURE 5. Transactivation by GAL4-CRP2 hybrids. A series of C-terminal truncations of CRP2 were fused to the GAL4 DNA-binding domain at the N terminus of CRP2. These proteins were tested in transactivation assays in HepG2 and L cells, using a GAL4-dependent promoter-CAT construct as the target gene.

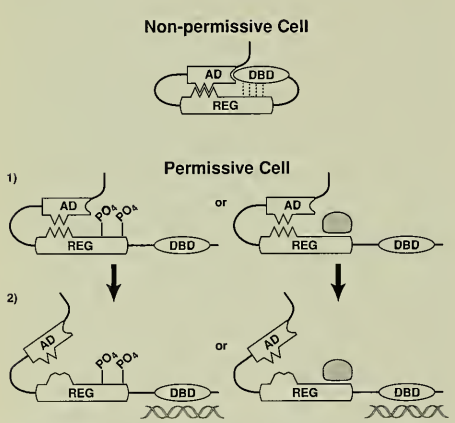


FIGURE 6. A model for the selective activity of CRP2 in permissive cell types. We propose that in nonpermissive cells, CRP2 adopts a tightly folded conformation (the "default state") that masks the activation and DNA-binding domains. This structure requires the integrity of the regulatory domain. In permissive cells, one of two events (phosphorylation or binding of another protein to the regulatory domain) causes partial unfolding of CRP2 (step 1). Further unfolding, or stabilization of the relaxed structure formed in step I, is proposed to occur when the protein binds to its cognate site through the bZIP DNA-binding domain (step 2). This second step fully exposes the activation domain, causing it to be maximally active. AD, activation domain; REG, regulatory domain; DBD, DNA-binding domain.

bZIP DNA-binding domain (Figure 6). We postulate that when complexed with the binding site, CRP2 undergoes a secondary structural transition that further exposes or potentiates the activation domain. Alternatively, DNA binding could simply stabilize the unfolding that is initiated by phosphorylation or protein binding in the regulatory region. In either case, our hypothesis is consistent with prior observations that upon binding to DNA, bZIP proteins isomerize such that the basic region becomes more α-helical. This localized structural change could alter the folded state of the entire protein and thereby influence the activity of the distally located activation domain.

Several potential phosphorylation sites in the serine/threonine-rich regulatory domain are possible targets for activating modifications. One of these (a MAP kinase site) has been shown to be phosphorylated in response to the *ras* signaling pathway, which stimulates CRP2 activity (Nakajima et al., Proc Natl Acad Sci USA 90:2207, 1993). We introduced amino acid substitutions into a cluster of these putative phosphorylation sites, including the MAP kinase site. None of these mutations diminished CRP2 activity in HepG2 cells (Figure 7) and, therefore, the mutated sites are probably not targets for modifications that are required to activate CRP2 in HepG2 cells. However, these variants have not yet been tested in L cells to determine whether the mutations constitutively inactivate the regulatory domain. We are currently investigating this possibility,

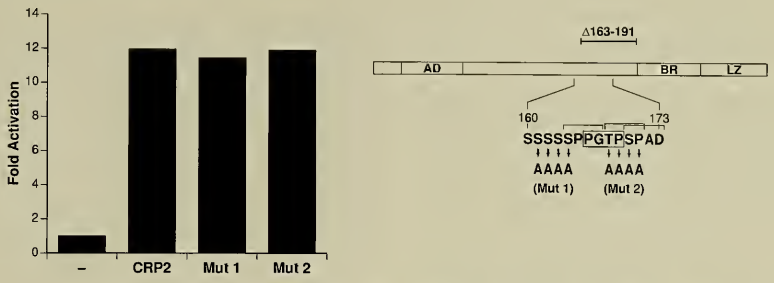


FIGURE 7. Amino acid substitutions affecting a cluster of potential phosphorylation sites in the regulatory region do not affect CRP2 activity in HepG2 cells. Two mutants (mut1 and mut2) were constructed that contain multiple amino acid substitutions in a serine/threonine-rich region of the regulatory domain. These sites lie within a deletion (Δ163-191) that inactivates the regulatory domain. The boxed sequence indicates a MAP kinase site that is phosphorylated in vivo in response to activated *ras* (Nakajima et al., Proc Natl Acad Sci USA 90:2207, 1993). The brackets denote other potential kinase sites that are affected by the mutations. The fold activation of a C/EBP-dependent reporter gene by each of the proteins in HepG2 cells is shown in the left panel. AD, activation domain; BR, basic region; LZ, leucine zipper.

and are testing other potential kinase sites in CRP2 for their involvement in cell-specific regulation. We also plan to perform phosphopeptide analysis of CRP2 expressed in HepG2 and L cells to identify sites that may exhibit differential phosphorylation in the two cell types.

Unlike CRP2, there is no evidence that the DNA-binding activity of C/EBP is regulated, even though C/EBP exhibits the same cell-type specificity as CRP2. However, a hybrid protein in which the GAL4 DNA-binding region is fused to the N terminus of C/EBP transactivated tenfold more efficiently in HepG2 cells than in L cells, regardless of whether a GAL4- or C/EBP-driven promoter was used as the target. The fact that C/EBP is fully cell specific when bound via the heterologous GAL4 DNA-binding domain implies that the activating region of C/EBP, and not its DNA-binding domain, is subject to cell-type regulation. In this respect, the mechanisms that generate cell specificity in C/EBP and CRP2 seem to be distinct.

Our studies show that CRP2 can be activated by a post-translational mechanism that functions in some cells but not others. Whether this differential activity results from actual cell-type differences or from some other property that varies among cell lines, such as their transformed/immortalized state, is not yet clear. We intend to address this question by measuring CRP2 activity in additional hepatic and nonhepatic cell lines. Furthermore, we hope to establish an in vivo assay system for CRP2 and C/EBP using transgenic mice, whereby we can assess their activity in a variety of normal cells and tissues. These experiments should clarify whether the liver-enriched CRP2 and C/EBP proteins are selectively functional in cell types such as hepatocytes.

Accumulating evidence suggests that CRP2 is an intricate regulatory protein that responds to multiple extracellular signals and interacts synergistically with several other transcription factors, including NF-κB, glucocorticoid receptor, and Sp1, to activate transcription. The inactive default state of CRP2 is probably critical for its rapid response to extracellular signals, which in two cases [interleukin-6 (IL-6) and phorbol ester] is known to entail the phosphorylation of specific serine residues. The response to these signals may involve a mechanism that is similar to the cell-specific activation of CRP2, where the inhibitory effects of the regulatory domain are probably reversed by protein modifications. We speculate that protein-protein interactions with other transcription factors might also activate CRP2 through structural transitions that cause functional domains to become exposed. In future studies, we will investigate the mechanisms by which CRP2 is activated by extracellular signals and interacts with other transcription factors. These experiments will rely on our knowledge of the domain structure of CRP2 and will be facilitated by the extensive set of CRP2 mutants that we now possess.

Regulation of CRP2 in Response to Cytokines and Its Role in Neuronal Differentiation

CRP2 expression is induced in a number of tissues by IL-6, interleukin-1, or bacterial lipopolysaccharide. In the myeloid lineage of the hematopoietic system, CRP2 expression correlates with the differentiated macrophage phenotype, which can be promoted by leukemia inhibitory factor (LIF) or IL-6. IL-6, which elicits the acute-phase response in hepatocytes, enhances CRP2 expression and activity in these cells. The induction of CRP2 activity by IL-6 (Nakajima et al., Proc Natl Acad Sci USA 90:2207, 1993) and phorbol ester (Trautwein et al., Nature 364:544, 1993) appears to involve protein phosphorylation events. CRP2 is also expressed in PC12 cells, a line derived from a tumor of rat adrenal chromaffin cells. In these cells, forskolin stimulates phosphorylation of CRP2 and increases the localization of the protein to the nucleus (Metz and Ziff, Genes Dev 5:1754, 1991). Collectively, these findings imply that CRP2 is an important component of the transcriptional response of cells to certain extracellular signals, particularly those that induce differentiation.

PC12 cells belong to the sympathoadrenal (SA) lineage, which derives from the neural crest and gives rise to sympathetic neurons and the neuroendocrine chromaffin cells of the adrenal medulla. Several factors such as basic fibroblast growth factor, nerve growth factor (NGF), ciliary neurotrophic factor (CNTF), LIF, and glucocorticoids promote differentiation of the SA lineage (reviewed by Patterson, Cell 62:1035, 1990). PC12 cells differentiate into cells resembling sympathetic neurons when treated with either NGF or IL-6 (Green and Tischler, Proc Natl Acad Sci USA, 73:2424, 1976; Satoh et al., Mol Cell Biol 8:3546, 1988). The NGF receptor, the *trk* proto-oncogene product, signals ligand binding via its tyrosine kinase activity (Kaplan et al., Nature 350:158, 1991). IL-6 elicits cellular responses via stimulation of the gp130 protein of its heterodimeric receptor, the signal-transducing activity of which is unknown. Thus, the same cellular response (neurite outgrowth) is elicited both by a tyrosine kinase-encoding receptor (in response to NGF) and a non-kinase–encoding receptor (in response to IL-6).

A large body of data exists on the signal transduction machinery of PC12 cells as well as on the genes regulated by different extracellular stimuli (reviewed by Haleguoa et al., Current Topics Micro Immunol 165:119, 1991). By comparison, few studies have focused on the direct regulation of transcription factors, which convert the extracellular signals into altered gene expression patterns. The fact that PC12 cells, like myeloid cells and hepatocytes, express CRP2 and contain a gp130 signaling pathway suggests that CRP2 mediates gene expression in the SA lineage in response to IL-6 signals. We therefore undertook studies to elucidate the activity and regulation of CRP2 in response to IL-6/gp130 and NGF/Trk mediated signaling pathways, and to investigate the potential role of CRP2 in the differentiation of SA cells.

Analysis of Transcriptional Activation by CRP2. In order to analyze whether CRP2 functions as a transcriptional activator in PC12 cells and whether its activity is influenced by cytokine receptor activation, we cotransfected PC12 cells with CRP2 expression plasmids and luciferase reporter constructs. In all the experiments described, we used a PC12 variant that we found to be most responsive to IL-6. Figure 8 shows data obtained with reporter constructs that contain either a segment of the chicken myelomonocytic growth factor (MGF) promoter, consisting of a TATA box and two CRP binding sites

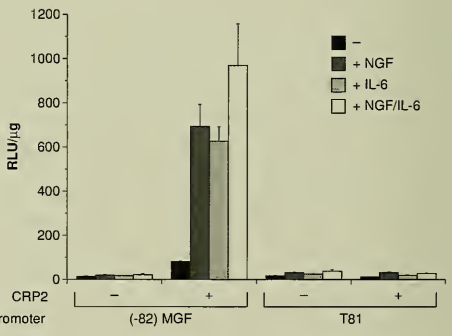


FIGURE 8. Enhancement of CRP2 transactivation by NGF and IL-6 in PC12 cells. PC12 cells were cotransfected with luciferase reporter constructs driven by either the (-82)MGF promoter or the T81 promoter (see text) in the presence or absence of a CRP2 expression plasmid. Cell extracts were prepared after 24 hours of treatment with 100 ng/ml of murine NGF and/or 100 ng/ml of human IL-6. Means and standard deviations of the luciferase activity (arbitrary units) per microgram of protein in duplicate transfections are shown. RLU, relative luciferase units.

[(-82)MGF], or the T81 promoter, consisting of a TATA box and one GC box (Nordeen, Biotechniques 6:454, 1988). When transfected into PC12 cells, both constructs yielded low levels of luciferase activity that were increased 1.5- to 3-fold in the presence of IL-6. Cotransfection of CRP2 enhanced expression of the (-82)MGF promoter, indicating that CRP2 functions as a transcriptional activator in PC12 cells. Exposing the cells to NGF and/or IL-6 increased luciferase expression 8- to 12-fold. Cotransfection of CRP2 had no effect on expression from the T81 promoter, which does not contain CRP binding sites. Furthermore, IL-6 or NGF did not alter T81 activity significantly in the presence or absence of CRP2. These data show that the differentiation-inducing cytokines NGF and IL-6 specifically enhance CRP2-mediated transcription. Because saturating concentrations of IL-6 and NGF were used in these experiments, the further enhancement observed with IL-6 and NGF together may indicate that the two factors activate CRP2 by at least partly different mechanisms.

In order to characterize the molecular mechanism of CRP2 stimulation by cytokines, we intend to analyze the response of CRP2 mutants (see previous section) to IL-6 and NGF receptor activation. We will also use PC12 variants that are impaired in specific signal transduction pathways in this assay. This approach may lead to some insight into the signal transduction pathway that connects receptor activation to CRP2 activation.

Analysis of CRP2 Expression. We have analyzed the expression of CRP2 in PC12 cells by Western blot analysis. Figure 9 shows that nuclear extracts of undifferentiated PC12 cells contain CRP2, as observed previously. We found that cytoplasmic fractions did not contain detectable levels of CRP2, and no other members of the C/EBP family were detectable in PC12 cells (data not shown). In differentiated cells, the expression level of CRP2 was altered depending on which inducer was used (Figure 9). IL-6 increased expression of CRP2, whereas NGF decreased CRP2 levels. Interestingly, in cells differentiating in the presence of both cytokines, the level of CRP2 appeared unchanged, suggesting that the positive effect of IL-6 on CRP2 expression is dominant to the suppressive effect of NGF. At the time of nuclear extract preparation, the cytokine-treated cells showed significant neurite outgrowth. Thus, the amount of CRP2 does not directly correlate with the differentiated or undifferentiated phenotype. However, activation of CRP2 expression appears to be IL-6 specific.

Analysis of CRP2 DNA-Binding Activity. We analyzed the DNA-binding activity of CRP2 in PC12 nuclear extracts by electrophoretic mobility shift assays. Figure 10 shows that a labeled oligonucleotide representing one of the CRP binding sites in the MGF promoter formed several complexes with nuclear proteins. Addition of CRP2-specific antiserum to the binding reaction led to supershifts of three of these complexes, showing that they contain CRP2. The identification of three complexes containing CRP2 suggests that CRP2 heterodimerizes with at least two different partners. The abundance of these complexes is enhanced in nuclear extracts of IL-6–treated cells, reduced in NGF-treated cells, and unchanged in cells treated with IL-6 plus NGF (Figure 10). None of the nonspecific DNA-binding complexes observed with the CRP site probe were significantly altered upon differentiation. These data show that CRP2 binding activity parallels its expression level, as detected by Western blot analysis (see Figure 9). Thus, the DNA-binding activity of CRP2 molecules appears unaltered in differentiating or undifferentiated PC12 cells.

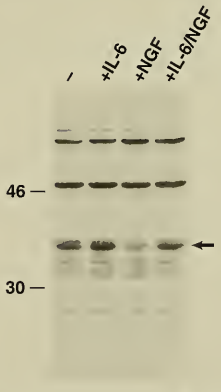


FIGURE 9. Expression of CRP2 in PC12 cells. Western blot analysis was performed using 50 µg of PC12 cell nuclear extract per lane. Cytokine treatments were for 4 days with 100 ng/ml of murine NGF and/or 100 ng/ml of human IL-6. The Western blot was probed with rabbit antiserum raised against an N-terminal peptide of CRP2. The arrow indicates the CRP2-specific signal. Molecular size markers (in kilodaltons) are indicated on the left.

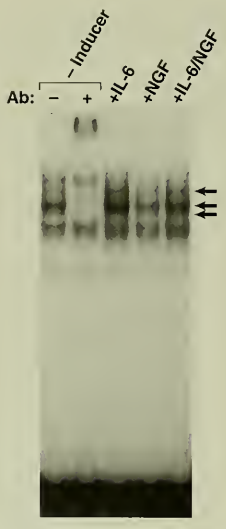


FIGURE 10. CRP2 binding activity in nuclear extracts of PC12 cells. Four micrograms of nuclear extract from undifferentiated or differentiated PC12 cells (see Figure 9) were incubated with a radioactively labeled oligonucleotide spanning the sequence of one of the CRP binding sites in the (-82)MGF promoter (5′-GGGCACTGTTGCCACATCGGTGCCC-3′) and analyzed by 5% polyacrylamide gel electrophoresis. Ab, binding reaction in the presence of CRP2-specific antiserum. Arrows indicate the CRP2-containing complexes.

The Role of CRP2 in Neuronal Differentiation. What is the purpose of activating CRP2 expression in response to IL-6 in PC12 cells? The signal-transducing component of the IL-6 receptor, gp130, is also a component of the receptors for the cytokines LIF and CNTF, which can induce cholinergic neuronal differentiation (reviewed by Hall and Rao, Trends Neurosci 15:35, 1992). IL-6 could therefore induce or modulate the phenotype of differentiating PC12 cells toward the cholinergic lineage, whereas NGF shunts differentiation along the noradrenergic lineage. CRP2 expression could be an important factor for activating the cholinergic transcriptional program. In order to test this hypothesis, we will analyze expression of specific markers for different neuronal phenotypes in PC12 cells differentiated by IL-6 or NGF and in cell lines overexpressing CRP2 (see below).

How do the data on CRP2 expression and DNA-binding activity conform with the data obtained by transient expression assays? Activation of CRP2-mediated transcription in cells exposed to IL-6 implies that not only is expression of endogenous CRP2 induced by IL-6, but that the activation potential of CRP2 also increases. This scenario parallels the observation made in hepatoma cells. NGF treatment also enhanced CRP2-dependent transcription, but the expression level of endogenous CRP2 was reduced. For the data from transient expression assays, cell extracts were prepared after 24 hours of cytokine treatment, whereas the expression levels of CRP2 shown in Figure 9 represent 4 days of exposure to cytokines. It is possible that CRP2 is activated by NGF specifically in early stages of differentiation and downregulated at a later stage. We are currently analyzing the kinetics of CRP2 expression and activation of reporter expression by IL-6 and NGF in PC12 cells.

It is important to note that transient expression assays examine the activity of exogenously expressed CRP2. Although the signal transduction pathway for NGF-induced activation of CRP2 may be present in the cell, it is not necessarily relevant for CRP2-regulated genes, as CRP2 expression itself is downregulated under these conditions. In order to analyze the role of CRP2 in PC12 cells more directly, we plan to overexpress CRP2 or inhibitors of CRP2 function, such as CRP2 antisense RNA or the CRP-specific inhibitor protein LIP (Descombes and Schibler, Cell 67:569, 1991), in stably transfected lines. We have made expression constructs for these genes with a tetracycline-regulated promoter and are currently selecting clones that have stably integrated these constructs. Subsequently, we will analyze whether their expression can promote, inhibit, or modulate the differentiated phenotype of PC12 cells in the presence or absence of differentiation inducers.

Cellular Growth Mechanisms Group

Deborah K. Morrison

Vaughn Cleghon
Jack Fabian
Karen Mathes

The objective of the Cellular Growth Mechanisms Group is to identify and study the proteins involved in regulating cell growth processes. We have focused our attention on the Raf family of serine/threonine kinases. The *raf* genes were originally identified as oncogenes, and activated Raf proteins have been found in a variety of human and rodent tumors. Previous work from our laboratory and others has demonstrated that the Raf family member Raf-1 plays an important role in the transmission of proliferative, developmental, and oncogenic signals. To gain a better understanding of how growth factors and oncogenic events activate the kinase activity of Raf-1, a major goal of our research has been to investigate the biochemical basis of Raf-1 regulation. In addition, to further elucidate the biological role of this family of proteins, we have initiated experiments to examine the involvement of *raf* gene products in mammalian, *Xenopus*, and *Drosophila* development. We hope that these studies will provide insight into how normal growth and developmental pathways function and how these processes are perturbed during oncogenesis.

Regulation of the Raf-1 Serine/Threonine Kinase

Identification of Raf-1 Phosphorylation Sites. Elucidation of the function and position of Raf-1 in signal transduction pathways has been the topic of many recent reports. Biochemical and genetic studies have demonstrated that one of the key pathways involved in the transmission of proliferative, developmental, and oncogenic signals from receptor tyrosine kinases (RTKs) to the nucleus involves the activation of $p21^{ras}$, Raf-1, and mitogen-activated protein kinase (MAPK, also known as ERK). Furthermore, this pathway functions in many cell types and in organisms as diverse as *Caenorhabditis elegans*, *Drosophila*, *Xenopus*, and mammals. In mammalian cells, activation of $p21^{ras}$ and Raf-1 by receptors that stimulate cellular tyrosine phosphorylation results in the sequential activation of the serine/threonine kinase activities of MEK (*M*APK kinase or *E*RK *K*inase, also known as MKK1), MAPK, and RSK. The activation of Raf-1 in many cases is dependent on the activity of $p21^{ras}$ and can be induced by expression of oncogenic *ras*, indicating that Raf-1 may function distally to $p21^{ras}$. In addition, $p21^{ras}$ enhances the kinase activity of Raf-1 when these proteins are coexpressed in the baculovirus/Sf9 system. However, for maximal activation of Raf-1 in this system, the presence of activated tyrosine kinases is also required. Since oncogenic forms of Raf-1 can phosphorylate and activate MEK, thereby activating MAPK and RSK, these kinases are thought to function downstream of Raf-1. In addition, immunoprecipitates of mitogen-stimulated Raf-1 are capable of phosphorylating and activating MEK in vitro. Raf-1 thus plays a central role in the transmission of growth signals. However, the way in which these signals are relayed to Raf-1 and how Raf-1 becomes activated are poorly understood.

A model that we have proposed for regulating Raf-1 activity involves phosphorylation events. In mammalian cells, reversible phosphorylation on serine, threonine, and tyrosine residues is a common mechanism used to regulate the function of proteins. Support for the model that

phosphorylation plays a role in regulating Raf-1 comes from the observation that the treatment of cells with various growth factors and mitogens results in the rapid hyperphosphorylation and activation of the Raf-1 kinase. In resting cells, Raf-1 is phosphorylated exclusively on serine residues. The treatment of cells with various growth factors and mitogens results in the rapid hyperphosphorylation of Raf-1. The de novo phosphorylation of Raf-1 in mitogen-stimulated cells occurs predominantly on serine residues, although increases in threonine and tyrosine phosphorylation have been observed. Concurrent with the hyperphosphorylation, the kinase activity of Raf-1 is enhanced in mitogen-treated cells. Therefore, the close correlation between Raf-1 phosphorylation and kinase activation suggests that phosphorylation may regulate the kinase activity. Determining the validity of this model requires the identification of the critical phosphorylation sites and mutational analysis of these sites. In collaboration with Terry Copeland (Special Program in Protein Chemistry) and Gisela Heidecker (National Cancer Institute, FCRDC), we initiated experiments to identify sites of Raf-1 phosphorylation.

By analysis of protein sequencing data, phosphorylation of synthetic peptides, and loss of phosphorylated peptides after site-directed mutagenesis, we determined that Ser43, Ser259, and Ser621 are three major serine residues of Raf-1 that are phosphorylated in vivo (Figure 1). In addition, we found that coexpression of Raf-1 with activated tyrosine kinases and $p21^{ras}$ in the baculovirus/Sf9 system results in the tyrosine phosphorylation of Raf-1 on Tyr340 and Tyr341 (Figure 1). Phosphorylation of these in vivo sites does not appear to be interdependent, since mutation of any one site does not affect the phosphorylation state of other sites. Ser43, Ser259, Ser621, and Tyr 340,341 also do not appear to be autophosphorylation sites, since Raf-1 proteins lacking kinase activity are phosphorylated on these sites in vivo when expressed under similar conditions as the wild-type Raf-1 protein. Furthermore, we have identified Thr268, located in the serine/threonine-rich CR2 region, as the major site of Raf-1 autophosphorylation in vitro. We previously identified an analogous threonine residue (Thr372) as the major in vitro autophosphorylation site of the B-Raf protein. The CR2 region is very highly conserved among the Raf family members, and residues within this domain may represent autophosphorylation sites of many of the Raf proteins. To date, we have not identified Thr268 as an in vivo site of Raf-1 phosphorylation; however, we cannot exclude that this might be a minor site phosphorylated in vivo following growth factor treatment of cells.

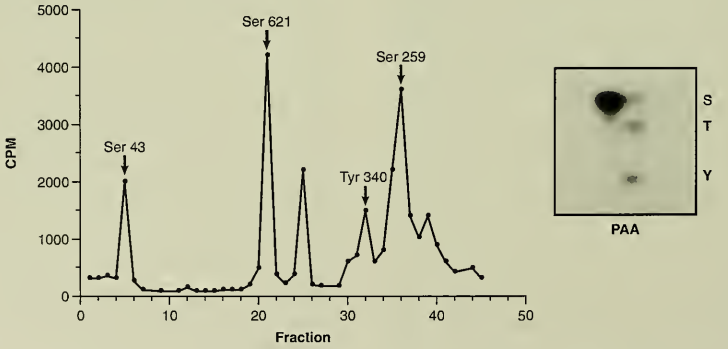


FIGURE 1. Identification of in vivo Raf-1 phosphorylation sites by reversed-phase high-pressure liquid chromatography (HPLC) analysis and phosphoamino acid analysis of in vivo ^{32}P-labeled Raf-1 phosphopeptides. In vivo-phosphorylated Raf-1 isolated from Sf9 cells coexpressing Raf-1, $p21^{ras}$, and activated $pp60^{src}$ was either subjected to phosphoamino acid analysis (PAA) or digested with trypsin. The tryptic phosphopeptides were then separated and eluted from a reversed-phase HPLC C_{18} column. The amount of ^{32}P radioactivity collected in each column fraction is shown. Phosphorylated residues found in fractions 5, 21, 32, and 36 correspond to Ser43, Ser621, Tyr340, and Ser259, respectively. S, phosphoserine; T, phosphothreonine; Y, phosphotyrosine.

Analysis of the Enzymatic and Biological Activity of the Raf-1 Phosphorylation Site Mutants. To determine the role of phosphorylation in the regulation of Raf-1 activity, we mutated the residues that are phosphorylated in vivo and examined the enzymatic and biological activity of Raf-1 proteins containing mutations at these sites. We analyzed the mutant proteins for enzymatic activity in protein kinase assays (Figure 2), for transforming potential in Balb/3T3 cells, and for the ability to induce the meiotic maturation of *Xenopus* oocytes (Figure 3). A summary of our findings is depicted in Figure 4.

The Ser43 in vivo phosphorylation site is located in the N-terminal regulatory domain upstream of the cysteine-rich CR1 region. This site was found to be phosphorylated

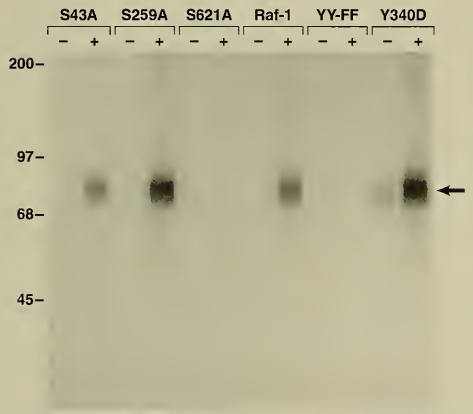


FIGURE 2. Analysis of the in vitro kinase activity of wild-type and mutant Raf-1 proteins. In vitro autokinase assays were performed using wild-type and mutant Raf-1 proteins immunoprecipitated from Sf9 cells either singly infected with baculoviruses expressing Raf-1 proteins (–) or triply infected with baculoviruses expressing Raf-1, $p21^{ras}$, and activated $pp60^{src}$ (+). Wild-type Raf-1 (Raf-1) and mutant Raf-1 proteins containing a serine to alanine substitution at amino acid position 43 (S43A), position 259 (S259A), or position 621 (S621A), a tyrosine to aspartic acid substitution at position 340 (Y340D), and tyrosine to phenylalanine substitutions at positions 340 and 341 (YY-FF) were examined. All samples were resolved by electrophoresis on 7.5% sodium dodecyl sulfate (SDS)-polyacrylamide gels, and the phosphoproteins were visualized by autoradiography. Migration of molecular size standards (in kilodaltons) is shown on the left.

both in untreated cells and in cells treated with platelet-derived growth factor (PDGF). Mutant proteins containing an alanine residue at position 43 (S43A/Raf-1) functioned as wild-type Raf-1 in all of our assays, indicating that mutation of this site has no effect on the biological or enzymatic activity of the Raf-1 protein. By sequence comparison of the Raf protein family, the Ser43 phosphorylation site appears to be present in all Raf-1 proteins, including those isolated from various species such as human, chicken, and *Xenopus*. However, the sequence of A-Raf, B-Raf, or *Drosophila* Raf proteins is divergent in this region and the Ser43 phosphorylation site is not found in these proteins. The residues surrounding the Ser43 site, RRASDDGK, constitute a suitable substrate for several serine/threonine-specific protein kinases, including protein kinase C (which has a specificity for the consensus sequence RxS/T) and cyclic AMP-dependent protein kinase (which has a specificity for the consensus sequence RRxS).

The Ser259 in vivo phosphorylation site is located in the serine/threonine-rich CR2 region. We found that Ser259 is a site of Raf-1 that is phosphorylated in response to PDGF treatment of cells. Surprisingly, mutant proteins that cannot be phosphorylated at position 259 were approximately twofold more active than was wild-type Raf-1 in the in vitro kinase assays. Furthermore, Raf-1 proteins containing an alanine residue at position 259 (S259A/Raf-1) were able to efficiently transform mammalian Balb/3T3 cells and were able to induce meiotic maturation of *Xenopus* oocytes. These findings suggest that phosphorylation at this site suppresses the activity of Raf-1. However, perturbation of the CR2 region by deletion or linker insertion activates the transforming potential of the Raf-1 protein, suggesting that this region plays a role in the negative regulation of Raf-1 activity. Therefore, any alteration in the CR2 region might result in Raf-1 activation. As a consequence, the change observed in the biological and enzymatic activities of the S259A mutant protein might not be related to phosphorylation events but, rather, to structural alterations (mutation of a serine residue to alanine) occurring in this highly conserved domain.

The Ser621 in vivo phosphorylation site is located at the end of the Raf-1 CR3 kinase domain. Like Ser43, this residue was phosphorylated in resting and PDGF-treated cells. In addition, Ser621 is a site phosphorylated in the N-terminal-truncated transforming Raf proteins (Δ-N′Raf).

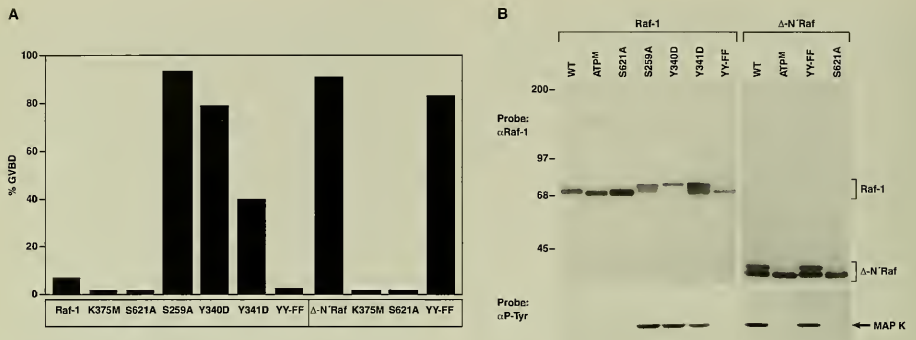
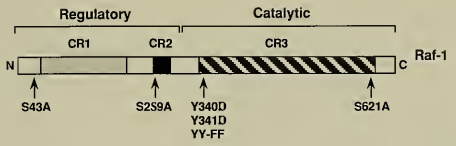


FIGURE 3. Expression of wild-type and mutant Raf-1 and Δ-N´Raf proteins in *Xenopus* oocytes. (A) Wild-type and mutant Raf-1 and Δ-N´Raf proteins were expressed in *Xenopus* stage VI oocytes and the induction of meiotic maturation was determined. At 18 to 24 hours after injection with RNA encoding Raf-1 and Δ-N´Raf, oocytes were scored for GVBD, as evidenced by the appearance of a white spot at the animal pole. (B) Anti–Raf-1 and antiphosphotyrosine Western blot analysis of Raf-1 and Δ-N´Raf proteins produced in *Xenopus* oocytes. Oocytes examined in this analysis were injected with RNA encoding full-length Raf-1 proteins [wild-type Raf-1 (WT), mutant Raf-1 containing a lysine to methionine substitution at amino acid position 375 (ATP^M), a serine to alanine substitution at positions 621 (S621A) and 259 (S259A), a tyrosine to aspartic acid substitution at positions 340 (Y340D) and 341 (Y341D), and tyrosine to phenylalanine substitutions at positions 340 and 341 (YY-FF)]; or truncated Δ-N´Raf proteins [wild-type Δ-N´Raf (WT), mutant Δ-N´Raf containing a lysine to methionine substitution at position 375 (ATP^M), tyrosine to phenylalanine substitutions at positions 340 and 341 (YY-FF), and a serine to alanine substitution at position 621 (S621A)].

Phosphorylation at this site appears to be important in the structural conformation of the catalytic domain, since mutation of this site inactivates the biological and enzymatic activity of the kinase. Interestingly, both the Ser259 and Ser621 phosphorylation sites are present in all Raf family members identified, including Raf-1, B-Raf, A-Raf, D-Raf, and *C. elegans* Raf. In addition, conserved residues include not only the sites of phosphorylation, but also the residues surrounding the phosphorylation sites. Analysis of these conserved residues indicates that RSxSxP may be a consensus sequence for phosphorylation of Raf-1 at Ser259 and Ser621.

Tyr340 and Tyr341, two adjacent tyrosine residues located in the kinase domain of Raf-1, are the in vivo sites of Raf-1 tyrosine phosphorylation. From protein sequencing data, Tyr340 was the major site of tyrosine phosphorylation, while Tyr341 was a minor phosphorylation site. However, if Tyr340 was changed to a negatively charged aspartic acid residue, Tyr341 became a major site of phosphorylation. Our findings indicate that substitution of a negatively charged amino acid for the Tyr340 or Tyr341 residue activated the in vitro and in vivo activity of Raf-1, presumably by mimicking constitutive phosphorylation of these sites. In Balb/3T3 cells, expression of either Y340D/Raf-1 or Y341D/Raf-1 mutant proteins resulted in morphological transformation. The extent of transformation, as measured by focus formation, induced by the Y340D mutant was equivalent to that of Δ-N´Raf. Interestingly, Y340D/Raf-1 is the first demonstration that a highly transforming Raf-1 protein can be generated by a single amino acid substitution. All other transforming Raf-1 proteins have been generated by more extensive mutations, such as N-terminal truncation, linker insertion, and fusion to other protein sequences. Expression of Y340D/Raf-1 was also able to induce germinal vesicle breakdown (GVBD), histone H1 kinase activity, and the activation of MAPK in *Xenopus* oocytes. In both systems, Y340D/Raf-1 was considerably more active than Y341D/Raf-1, reflecting the higher relative enzymatic activity of the Y340D mutant. In contrast, replacing the Tyr340 and Tyr341 residues with noncharged amino acids that were unable to be phosphorylated (YY-FF/Raf-1) generated a Raf-1 protein that could not be enzymatically activated by tyrosine kinases or $p21^{ras}$.



Mutant	Phenotype
1) S43A	-Wild-type
2) S259A	-Transforms Balb/3T3 cells -Induces GVBD in *Xenopus* oocytes -Enhanced kinase activity
3) S621A	-Dominant-inhibitory -Kinase defective
4) Y340D	-Transforms Balb/3T3 cells -Induces GVBD in *Xenopus* oocytes -Constitutively active kinase
5) Y341D	-Y341D activity <Y340D activity
6) YY-FF	-Dominant-inhibitory -Functional catalytic domain -Kinase activity not activatable

FIGURE 4. Summary of Raf-1 phosphorylation site mutant proteins. A schematic depiction of the Raf-1 protein is shown. The three regions highly conserved among the Raf family members are also depicted: CR1 is rich in cysteine residues; CR2 is rich in serine/threonine residues; and CR3 is the kinase domain.

Furthermore, expression of the YY-FF/Raf-1 in *Xenopus* oocytes prevented the meiotic maturation induced by signaling components of the RTK pathway.

The position of the tyrosine residues in the Raf-1 kinase suggests that phosphorylation at these sites alters the conformation of the protein such that the kinase domain is released from the suppression exerted by the N-terminal regulatory domain. This model is supported by the observation that if tyrosine phosphorylation was prevented, the enzymatic activity of full-length YY-FF/Raf-1 failed to be induced by tyrosine kinases. However, the catalytic activity of truncated YY-FF/Δ-N´Raf was constitutively activated and unaffected by mutation at the tyrosine residues, demonstrating that the tyrosine to phenylalanine mutations did not impair kinase activity. Furthermore, expression of YY-FF/Raf-1 in oocytes inhibited the meiotic maturation induced by the Tpr-Met tyrosine kinase and H-ras^{V12}, while truncated YY-FF/Δ-N´Raf stimulated meiotic maturation. These results indicate that mutation of Tyr340 and Tyr341 to phenylalanine residues does not prevent the interaction of the kinase domain of Δ-N´Raf with downstream targets. Rather, the functional defect of the YY-FF/Raf-1 mutant appears to be caused by constitutive suppression of the kinase domain by the N-terminal regulatory domain, thereby inhibiting the interactions of this protein with its substrates. As a result, the dominant inhibitory effect observed for YY-FF/Raf-1 may be caused by the binding of upstream Raf activators, preventing the activation of the endogenous *Xenopus* Raf. These activators may include p21^{ras}, which has recently been shown to bind to Raf-1 in vitro.

Sequence alignment of the catalytic domains of Raf proteins reveals that all Raf-1 and A-Raf proteins have tyrosine residues at the positions analogous to Tyr340 and Tyr341, while B-Raf, *Drosophila* Raf, and *C. elegans* Raf have negatively charged residues at these positions. Since substitution of a charged amino acid at Tyr340 or Tyr341 activates Raf-1 catalytic activity, the Raf proteins encoding charged residues at these sites may be partially activated. Alternatively, the mechanism of the regulation of B-Raf, *Drosophila* Raf, and *C. elegans* Raf activity may differ from that of Raf-1 and A-Raf. Interestingly, the N-terminal domains of B-Raf, *Drosophila* Raf, and *C. elegans* Raf are much larger than those of Raf-1 and A-Raf. The additional sequences may encode a novel regulatory domain.

Examination of the Role of Raf and MAPK in the Meiotic Maturation of *Xenopus* Oocytes

Since Raf-1 and MAPK play an essential role in transducing proliferative signals in somatic cells, we initiated experiments in collaboration with Ira Daar (National Cancer Institute, FCRDC) to examine the requirement of Raf and MAPK function during the meiotic maturation of *Xenopus* oocytes. The meiotic maturation of *Xenopus* oocytes has been extensively used as a model system to study cell cycle regulation. Oocyte maturation can be induced by several events, including treatment with progesterone or expression of activated forms of p21^{ras} or RTK. Consistent with previous studies, we found that in progesterone-treated oocytes, MAPK underwent several readily observable physical changes at the onset of GVBD. These changes included tyrosine phosphorylation, an

alteration of electrophoretic mobility, and activation of its serine/threonine kinase activity (Figure 5). Concurrent with MAPK tyrosine phosphorylation and at a time closely associated with activation of maturation promoting factor (MPF), we observed a shift in the electrophoretic mobility of *Xenopus* Raf (Figure 5). The activation of MAPK stimulated by progesterone and proteins associated with RTK signaling pathways [EGF receptor, activated Met (hepatocyte growth factor) receptor, $p21^{v\text{-}ras}$] could be blocked by expression of a kinase-defective Raf mutant, S621A/Raf-1 (Figure 6). This result demonstrates that MAPK activation during the maturation of *Xenopus* oocytes is dependent on the activity of Raf and is consistent with observations of Raf function in mammalian somatic cells. Interestingly, expression of S621A/Raf-1 in oocytes completely blocked MPF activation and GVBD induced by the RTK pathway, but only delayed progesterone-stimulated GVBD and progression of maturation to meiosis II (Figure 7).

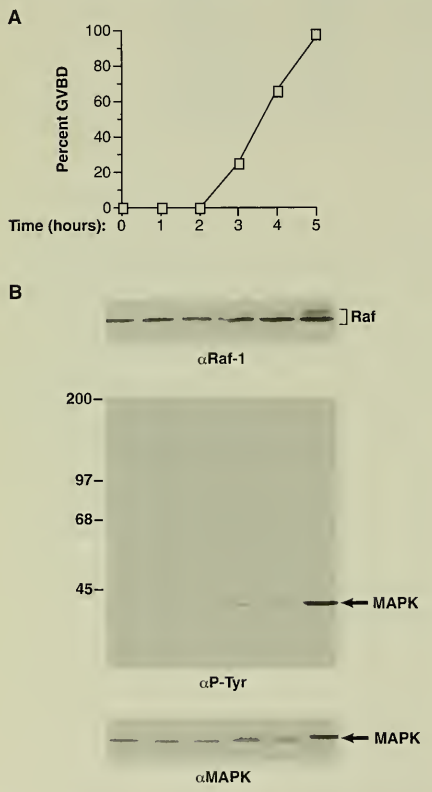


FIGURE 5. Progesterone-induced GVBD is coincident with *Xenopus* Raf and MAPK activation. (A) Eighty oocytes were treated with progesterone (2 μg/ml), and the percentage of oocytes undergoing GVBD was determined as a function of time. At the indicated time, 5 oocytes were arbitrarily collected and lysed. (B) Samples were resolved by electrophoresis on a 7.5% SDS-polyacrylamide gel, using 2.0 oocyte equivalents of lysates per lane, and examined by Western blot analysis using specific antibodies for Raf (αRaf), phosphotyrosine (αP-Tyr), or MAPK (α-MAPK). Migration of molecular size standards (in kilodaltons) is shown on the left.

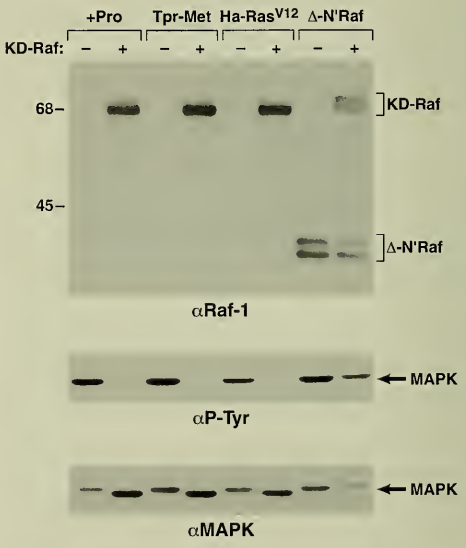


FIGURE 6. Inhibition of MAPK activation by kinase-defective Raf-1 (S621/Raf-1, or KD-Raf). Oocytes preinjected with S621A/Raf-1 (+) or not preinjected (–) were either treated with progesterone (+Pro) or microinjected with the capped transcripts (30 ng/oocyte) encoding Tpr-Met, Ha-ras^{V12}, or Δ-N′Raf. After 12 hours, oocytes were collected and lysates were prepared. Samples were resolved by electrophoresis on an 8.0% SDS-polyacrylamide gel, using 2.5 oocyte equivalents of lysate per lane, and examined by Western blot analysis using specific antibodies for Raf (αRaf-1), phosphotyrosine (αP-Tyr), or MAPK (αMAPK). Migration of molecular size standards (in kilodaltons) is shown on the left.

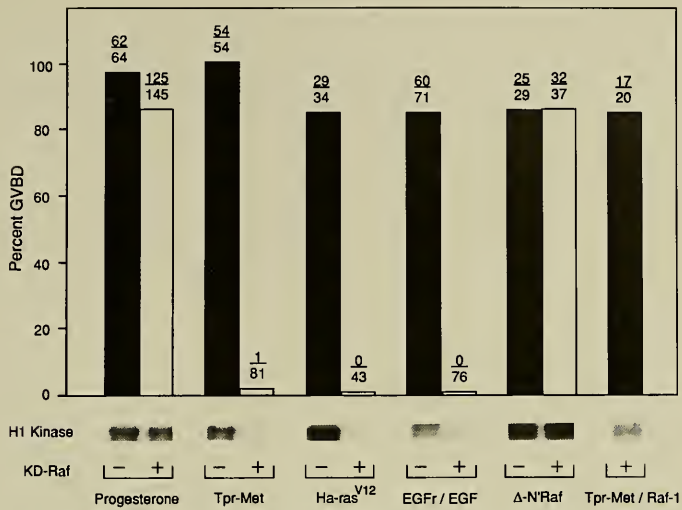


FIGURE 7. Expression of kinase-defective Raf-1 (S621/Raf-1, KD-Raf) inhibits GVBD induced by signal transducers of the RTK pathway. Oocytes preinjected with S621A/Raf-1 (+) or not preinjected (–) were either treated with progesterone or microinjected with the capped transcripts (30 ng/oocyte) encoding Tpr-Met, Tpr-Met and wild-type Raf-1 (Tpr-Met/Raf-1), Ha-ras^{V12}, or Δ-N′Raf. In the case of the EGF receptor (EGFr), EGFr RNA was injected 24 hours prior to S621A/Raf-1 RNA injection (+), or not preinjected (–), then stimulated with EGF (0.1 μg/ml). GVBD was examined 12 to 18 hours later. The ratio of the number of oocytes with GVBD to the total number injected is displayed above each bar. Histone H1 kinase assays were performed on extracts from 5 injected oocytes and the autoradiograph is displayed below each bar.

Furthermore, we found that the activation of MAPK in oocytes requires the expression of the Mos proto-oncogene product, pp39^{mos}. This result differs from those observed in somatic cells, where the activation of MAPK occurs in the absence of Mos expression.

In summary, we conclude that MAPK functions downstream of Raf during the maturation of *Xenopus* oocytes. In addition, our data suggest that Raf and MAPK are common intermediates for both tyrosine kinase and progesterone signaling pathways involved in oocyte maturation. However, while Raf is an essential signaling component of the tyrosine kinase pathway, its activities might only expedite cell cycle progression induced by progesterone. These findings further indicate that RTKs and progesterone participate in distinct yet overlapping signaling pathways resulting in the activation of MPF.

Eukaryotic Signal Transduction Group

David R. Kaplan

Neil R. Michaud
Stuart J. Rabin
Kyoichi Shibuya
Robert M. Stephens
Deborah Wancio
Sung-Il Yang

The objective of the Eukaryotic Signal Transduction Group is to understand the mechanism of signal transduction during mitogenesis, oncogenesis, and cell differentiation. As a model system, we have focused on signal transduction by nerve growth factor (NGF), a neurotrophic factor required for the survival and differentiation of specific populations of neurons. Three years ago, in collaboration with Dionisio Martin-Zanca and Luis Parada (Molecular Embryology Section, Mammalian Genetics Laboratory) and Barbara Hempstead and Moses Chao (Cornell University Medical School), we identified the Trk tyrosine kinase as the receptor for this factor. In neuronal cells, the activation of Trk induces cell differentiation and the cessation of cell proliferation. This observation has enabled us to examine the signal transduction pathways used to stop cell growth. By identifying and characterizing the proteins in cells that mediate the antimitogenic activity of Trk, we will gain a more cogent understanding of the mechanisms of cellular growth control. In addition, the investigation of Trk expression, activity, and function during nervous system development will provide insights into how neurotrophins control the proliferation and differentiation of neurons. Finally, we have identified several agents that enhance the expression or antiproliferative activity of Trk. These drugs may aid in the development of novel approaches to the treatment of human cancer. Moreover, since Trk and the members of the Trk family have important roles in regulating the development, survival, and regeneration of neurons, the ligands and agents that stimulate Trk activity are ideal candidates for treating degenerative nervous system diseases such as Alzheimer's disease.

While the signal transduction pathways used by tyrosine kinases regulating cell proliferation have been extensively studied, little is known about signal transduction by tyrosine kinases mediating cell differentiation. Neuronal cells provide an ideal system for the study of signal transduction during differentiation. Several neurotrophic factors bind receptor tyrosine kinases (RTK), including the NGF family of neurotrophins and fibroblast growth factor (FGF). The identification of the intracellular targets of neurotrophin receptors may provide insights into the mechanism of cell differentiation. More importantly, the study of signal transduction pathways in neuronal cells may provide a better understanding of cell growth control. During neuronal development, RTK must act in a single neuron to promote the opposing processes of cell differentiation and cell growth. For example, NGF causes PC12 cells to stop dividing and differentiate. However, the growth factors insulin and epidermal growth factor (EGF) are mitogenic for this cell line. Thus, the distinctive effects of different RTK may be studied in the same cell. The comparison of mitogenic and differentiative pathways could identify cellular proteins that determine the specificity of growth and differentiation factors.

In the past three years, we have focused on identifying all of the intracellular targets of NGF-mediated tyrosine kinase activity. In collaboration with several groups, including Stephen Soltoff and Lewis Cantley (Harvard Medical School), Monica Vetter and J. Michael Bishop (University of California, San Francisco), Tohru Kamata (Program Resources, Inc./DynCorp, FCRDC), and Deborah Morrison (Cellular Growth Mechanisms Group), we identified

phospholipase C-γ1 (PLC-γ1), phosphatidyl inositol-3 (PI-3) kinase, and SHC (Src homologous and collagen) as the first direct substrates of Trk. In addition, our collaborative efforts led to the identification of several novel targets of NGF, including GTPase-activating protein (GAP) and the ras guanine nucleotide exchange proteins (rasGEP) that are involved in regulating $p21^{ras}$ activity, and B-Raf, a serine/threonine kinase and proto-oncogene product that is highly expressed in the brain. A summary of our efforts and those of other investigators to identify targets of NGF-induced activity in PC12 cells is provided in Figure 1.

None of these targets has been demonstrated to be specifically activated by NGF and not by mitogens such as EGF. Thus, we postulated that one or more proteins must therefore be activated by Trk to initiate cell differentiation rather than mitogenesis. We initiated studies to identify novel substrates of Trk and Trk family members. These studies resulted in the identification of one such protein, SNT.

SNT: A Differentiation-Specific Target of Neurotrophic Factor-Induced Tyrosine Kinase Activity in Neurons and PC12 Cells

To identify proteins that act in differentiative rather than mitogenic signal transduction pathways, we examined the tyrosine-phosphorylated proteins in PC12 cells and primary neurons treated with neurotrophic factors or mitogens that stimulate cellular tyrosine kinase activity. Since the antimitogenic properties of neurotrophic factors may be mediated by proteins that interact with cell cycle regulators, we analyzed the association of tyrosine-phosphorylated proteins with an affinity chromatography reagent, $p13^{suc1}$-agarose, that binds $p34^{cdc2}$ and $p33^{cdk2}$. As a serine/threonine kinase that controls entry into mitosis, $p34^{cdc2}$ interacts with at least two proteins, cyclin and $p13^{suc1}$, that are required for $p34^{cdc2}$ activity. While both positive and negative effects on cell division have been attributed to $p13^{suc1}$, its true role is unknown.

Identification of SNT. Using $p13^{suc1}$-agarose affinity chromatography, we identified a 78- to 90-kilodalton species in PC12 cells that was rapidly phosphorylated on tyrosine following treatment with the differentiation factors NGF and FGF, but not with the mitogens EGF or insulin. This species, called SNT (*s*uc-associated *n*eurotrophic factor-induced *t*yrosine-phosphorylated target), was also phosphorylated on tyrosine in primary rat cortical neurons treated with the neurotrophic factors neurotrophin-3 (NT-3), brain-derived neurotrophic factor (BDNF), and FGF, but not EGF (Figure 2). SNT was phosphorylated within 1 minute of NGF addition to cells, indicating that the phosphorylation of this protein is one of the earliest responses of cells to neurotrophic factor treatment (Figure 3A). The dose of NGF required to induce half-maximal tyrosine phosphorylation of SNT

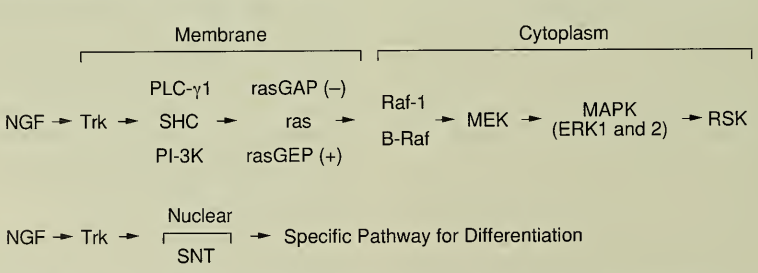


FIGURE 1. Trk signal transduction pathways. NGF binding to Trk stimulates the rapid tyrosine phosphorylation or activation of many signal-transducing proteins. Activated Trk receptors bind to PLC-γ1, PI-3 kinase, and SHC. SHC activates $p21^{ras}$ activity by binding to Grb2 and rasGEP. Ras is positively regulated by rasGEP and negatively regulated by rasGAP. Ras may then activate several serine/threonine kinases, including Raf-1, B-Raf, MEK [mitogen-activated protein kinase (MAPK) kinase], and RSK. MAPK and RSK may be involved in activating gene transcriptional events. All of the proteins in this pathway appear to function in cells treated with NGF or with the mitogen EGF. Trk also mediates the tyrosine phosphorylation of SNT, which localizes to nuclear fractions of cells. SNT is a specific target of neurotrophic factors and not mitogens for neurons and PC12 cells.

was between 1 and 10 ng/ml, indicating that the phosphorylation event occurred at physiologically relevant concentrations of NGF (Figure 3B).

Each of the factors assayed in the experiments described above is known to stimulate protein tyrosine phosphorylation in responsive cells. To determine whether the mitogens and neurotrophic factors induce similar tyrosine phosphorylation responses (with the exception of SNT), we assessed these responses in PC12 cells treated for 5 minutes with NGF, FGF, or EGF. Examination of cell lysates probed with phosphotyrosine antibody indicated that all three factors stimulated the tyrosine phosphorylation of both common and unique proteins (Figure 4A). Two of these proteins, PLC-γ1 and ERK1, were phosphorylated on tyrosine in PC12 cells treated with NGF, FGF, or EGF (Figure 4B and C). Thus, while several proteins are phosphorylated on tyrosine in response to neurotrophic factor and mitogen treatment of neurons and PC12 cells, only neurotrophic factors induce SNT tyrosine phosphorylation.

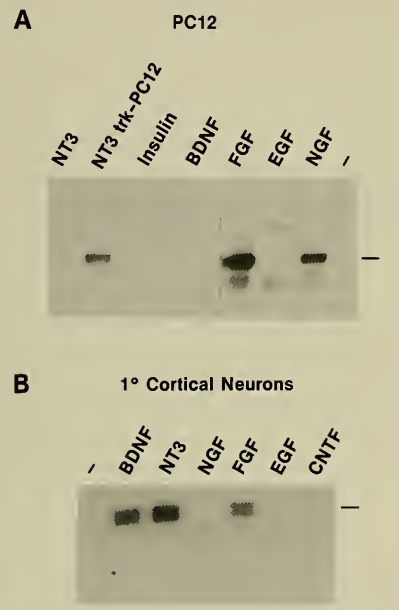


FIGURE 2. Tyrosine phosphorylation of SNT in PC12 cells and primary rat cortical neurons treated with neurotrophic factors and mitogens. Cells were treated with factors for 5 minutes and lysed, and cell lysates were precipitated with $p13^{suc1}$-agarose. Proteins were transferred to nitrocellulose and probed with phosphotyrosine antibody. (A) Growth factor specificity of SNT tyrosine phosphorylation in PC12 cells. Cells were treated with factors exhibiting neurotrophic activity (NGF, FGF, NT-3) or mitogenic activity (EGF, insulin) or with BDNF. PC12 cells lack receptors for BDNF and respond to NT-3 only when Trk is overexpressed in these cells (NT3 trk-PC12). (B) Growth factor specificity of SNT tyrosine phosphorylation in primary rat cortical neurons. Cultures were treated with factors exhibiting neurotrophic activity (BDNF, NT-3, FGF) or mitogenic activity (EGF) or with NGF and ciliary neurotrophic factor (CNTF). These cells lack receptors for NGF and CNTF.

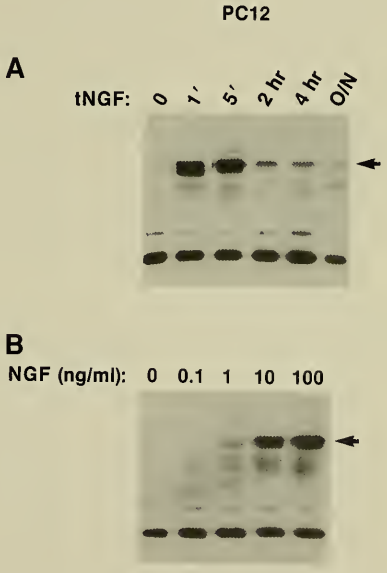


FIGURE 3. Time course and dose-response of SNT tyrosine phosphorylation in PC12 cells. Cells were treated with NGF and lysed, and cell lysates were precipitated with $p13^{suc1}$-agarose. Proteins were transferred to nitrocellulose and probed with phosphotyrosine antibody. (A) Time course of SNT tyrosine phosphorylation in PC12 cells treated for 5 minutes with NGF. The position of SNT is indicated with an arrow. (B) Dose-response of SNT tyrosine phosphorylation.

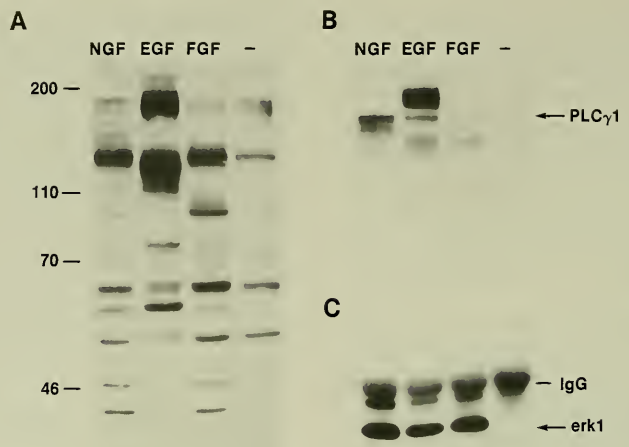


FIGURE 4. Growth factor-induced tyrosine phosphorylation of cellular proteins in PC12 cells. Cells were mock treated or treated with factors for 5 minutes. (A) Protein tyrosine phosphorylation in PC12 cells treated with NGF, EGF, or FGF. Cell lysates were probed with phosphotyrosine antibody. (B) Tyrosine phosphorylation of PLC-γ1. Growth factor-treated cells were lysed and the lysates were immunoprecipitated with PLC-γ1 antibody. PLC-γ1 immunoprecipitates were probed with phosphotyrosine antibody. The broad band above PLC-γ1 in the EGF lane is most likely the EGF receptor. The diffuse band below PLC-γ1 in the NGF lane is Trk. (C) Tyrosine phosphorylation of ERK1. Growth factor-treated cells were lysed and the lysates were immunoprecipitated with ERK1 antibody. ERK1 immunoprecipitates were probed with phosphotyrosine antibody. The positions of ERK1 and IgG are indicated.

NGF Induces SNT Tyrosine Phosphorylation While Acting as a Differentiation Factor and Not as a Mitogen. The PC12 variant cell line U7 responds to NGF by differentiating or proliferating based on the concentration of serum in the medium. In serum-free medium, NGF induced the differentiation of U7 cells, while in medium with 1.5% serum, NGF stimulated cell proliferation. We assayed SNT tyrosine phosphorylation in NGF-treated U7 cells in different serum concentrations. NGF induced significant SNT tyrosine phosphorylation only in serum-free medium, in which this factor promotes differentiation. Small amounts of tyrosine-phosphorylated SNT were evident in NGF-treated U7 cells grown in medium containing 1.5% serum. NGF stimulated the tyrosine phosphorylation of ERK1 both in serum-free medium and in medium containing 1.5% serum, indicating that the NGF signal transduction machinery was functioning in cells grown in both serum conditions.

SNT Is Localized in Nuclear Fractions of Cells. Determination of the intracellular location of SNT may suggest a possible role for this polypeptide. Vaughn Cleghon (Cellular Growth Mechanisms Group) fractionated PC12 cells treated for 1 or 5 minutes with NGF into cytosolic and nuclear fractions and analyzed the location of tyrosine-phosphorylated SNT. Most of the tyrosine-phosphorylated SNT was evident in the nuclear fraction (Figure 5A). The effectiveness of the fractionation procedure was assessed by monitoring the intracellular location of several previously characterized proteins. The B-Raf protein was found almost exclusively in the cytosolic fraction, whereas the nuclear transcription factor c-Jun was localized primarily in the nuclear fraction (Figure 5B). Tyrosine-phosphorylated SNT is thus primarily localized in nuclear fractions of NGF-treated PC12 cells.

We identified several other interesting properties of SNT. First, SNT was phosphorylated in vitro on serine, threonine,

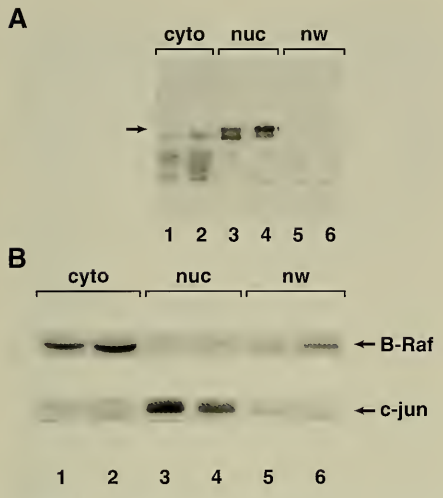


FIGURE 5. Association of tyrosine-phosphorylated SNT with nuclear fractions. PC12 cells were treated with NGF for 1 minute (lanes 1, 3, and 5) or 5 minutes (lanes 2, 4, and 6) and fractionated into cytoplasmic (cyto), nuclear (nuc), or nuclear wash (nw) fractions. (A) Samples were precipitated with $p13^{suc1}$-agarose, and the presence of SNT was assessed by probing with phosphotyrosine antibody. The position of SNT is indicated with an arrow. (B) Aliquots of cell fractions probed with an antibody to the cytoplasmic protein B-Raf (top panel) or with an antibody to the nuclear protein c-Jun (bottom panel).

and tyrosine in $p13^{suc1}$-agarose precipitates from NGF-treated PC12 cells, indicating that this protein may be a substrate of kinase activities associated with $p13^{suc1}$/$p34^{cdc2}$/$p33^{cdk2}$ complexes. Secondly, in PC12 cells, the NGF-stimulated tyrosine phosphorylation of SNT was dependent on the levels of Trk tyrosine kinase activity and was constitutively induced by expression of $pp60^{v\text{-}src}$. However, $p21^{ras}$ was not required for SNT tyrosine phosphorylation, suggesting that SNT may function downstream of Trk and $pp60^{v\text{-}src}$ but in a parallel pathway to $p21^{ras}$. SNT is the first identified specific target of differentiation factor-induced tyrosine kinase activity in neuronal cells. Since one of the effects of NGF on PC12 cells is a cessation of cell division, the association of SNT with the cell cycle protein $p13^{suc1}$ raises the possibility that SNT is involved in mediating the antimitogenic properties of NGF.

To precisely determine the role of the SNT protein, Neil Michaud, Kyoichi Shibuya, and Terry Copeland (Special Program in Protein Chemistry) will identify the amino acid sequence and isolate the molecular clone of SNT.

In addition to examining the signal transduction pathways used by Trk to initiate neuronal development, we have investigated several other aspects of Trk function. Last year, we determined the role of Trk tyrosine kinase activity in cultured cells. We found that in PC12 cells, Trk regulates the timing of NGF-induced neuronal differentiation. Robert Stephens has performed an extensive mutational analysis of Trk. His results, in collaboration with Lloyd Greene (Columbia University), resulted in the identification of the sequences of Trk that are responsible for interacting with several intracellular proteins involved in signal transduction and that mediate many of the biological responses to NGF. With Beat Knusel and Franz Hefti (University of Southern California), we assayed the expression and neurotrophin response pattern of Trk family members during development. Our results suggest that the different Trk proteins and neurotrophins play distinct roles during neuronal development and in the adult. Finally, with Carol Thiele (National Cancer Institute), we identified the morphogen retinoic acid (RA) as a powerful inducer of the expression of the Trk family member TrkB. These later data will be presented in the following section.

RA Induces TrkB, a Receptor that Mediates the Biological Responsiveness of Neuroblastoma Cells to Neurotrophins

The survival, maturation, and function of neurons is regulated by neurotrophic factors. The neurotrophins, which include NGF, BDNF, NT-3, and neurotrophin-4/5 (NT-4/5), exhibit structural similarities and show unique and overlapping temporal and spatial expression patterns. The effects of neurotrophins on neuronal cells are mediated by the cellular expression of specific Trk receptors. The Trk family of protein tyrosine kinase receptors (Trk, TrkB, and TrkC) encode both full-length proteins (M_r 130,000 to 145,000) and truncated variants that lack the intracellular tyrosine kinase domain. Trk is activated predominantly by NGF, whereas TrkB is activated by BDNF, NT-3, and NT-4/5, and TrkC is activated by NT-3. While several NGF-responsive, Trk-expressing neuronal cell lines

have been described, the analysis of differentiation induced by the neurotrophins has been inhibited by the lack of cell lines endogenously expressing TrkB and TrkC. We found that cell lines derived from human neuroblastomas (NB) have the potential to express TrkB.

Human NB is a model system for studying differentiation of cells of the sympathoadrenal lineage. NB are tumors of neural crest origin in which alterations that affect normal development may be involved in tumorigenesis. NB have been observed to spontaneously regress in vivo to benign ganglioneuromas, and many NB cell lines can be induced to differentiate by several agents including RA and NGF. In many NB cell lines, RA induces the arrest of NB cell growth in the G_1 phase of the cell cycle and induces differentiation into cells that are biochemically, ultrastructurally, and electrophysiologically similar to neurons. However, the molecular mechanisms mediating neuronal differentiation in NB cells are not known.

Several studies suggest that RA plays a role in neuronal development. RA alters the development of nervous system tissues in the chick and in *Xenopus laevis*, and retinoids and RA-binding proteins are localized in the developing chicken, mouse, and amphibian spinal cord and central nervous system. RA has trophic effects on embryonic rat and amphibian spinal cord neurons, including enhancements of neuronal survival and neurite outgrowth. RA also induces NGF responsiveness in several cell lines. The induction of neurotrophin responsiveness by RA suggested to us that this agent could stimulate the expression of neurotrophin receptors.

In collaboration with Dr. Thiele, we examined the mechanism of neuronal differentiation of NB cells by RA. We found that induction of the TrkB neurotrophin receptor by RA mediates biological responsiveness to the TrkB ligands BDNF, NT-3, and NT-4/5 and the differentiation of human NB cells.

RA and TrkB Ligands Induce the Differentiation of SY5Y NB Cells. When human SY5Y NB cells were treated with NGF, BDNF, or RA alone, no significant difference in the morphology of the cells was noted (Figure 6, panels 1 to 3A, 1 to 2B). The cells were more substrate adherent yet exhibited only limited neurite outgrowth responses. However, cells treated with both RA and BDNF exhibited a dramatic change in morphology characterized by an extensive network of neuritic processes involving the majority of the cells (Figure 6, panel 3B). Neurite extension was increased in cultures treated with RA and NT-3, although to a more limited extent than was seen in cultures treated with BDNF and RA. The striking morphologic change noted in SY5Y cells treated with both RA and BDNF (and to a lesser degree NT-3) and the known specificity of BDNF for TrkB receptors suggested that RA induced the expression of TrkB in SY5Y cells.

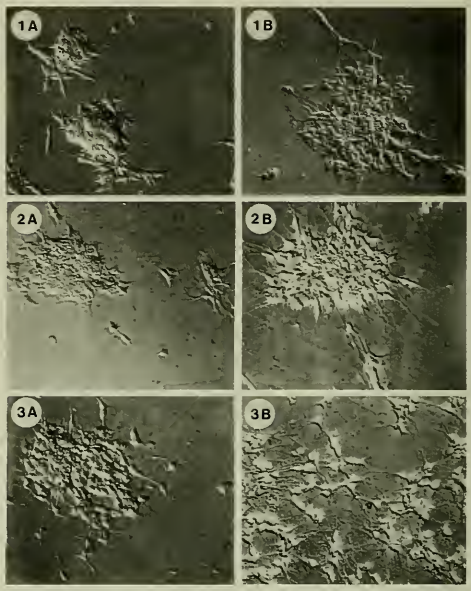


FIGURE 6. Differentiation of SY5Y NB cells by RA and BDNF. Cells were treated with control solvent (A) or with RA (1 μM) in solvent (B) in the absence of neurotrophin (panel 1) or in the presence of NGF (panel 2) or BDNF (panel 3). RA and neurotrophin treatment was administered for 6 days.

RA Stimulates the Appearance of TrkB- and Neurotrophin-Regulated Tyrosine Phosphorylation in SY5Y Cells. To assess whether RA induced the expression of TrkB, we performed immunoblot experiments using an antibody that specifically recognizes TrkB. Increases in TrkB levels were evident in SY5Y cells treated with RA for 4 days, with maximal levels at 8 days (Figure 7A, panel 1). RA also stimulated the appearance of a truncated 95-kilodalton TrkB

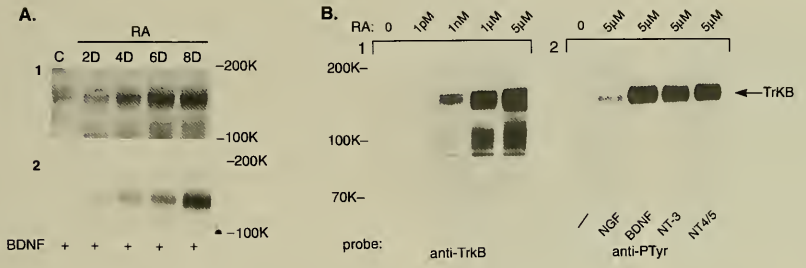


FIGURE 7. RA-induced appearance of TrkB protein and neurotrophin-induced TrkB tyrosine phosphorylation in SY5Y cells. (A) Stimulation of TrkB protein expression and BDNF-induced tyrosine phosphorylation in RA-treated cells. Control cells (C) or cells treated with RA for the indicated times (in days) were stimulated with BDNF for 5 minutes, and TrkB proteins were immunoprecipitated with Trk antibody. The immunoprecipitates were analyzed by probing with antibody to TrkB (panel 1) or phosphotyrosine (panel 2). (B) Dose response of RA induction of TrkB (panel 1). Lysates from cells treated with the indicated concentrations of RA for 8 days were precipitated with wheat germ lectin-agarose and probed with TrkB antibody. In panel 2, cells treated for 8 days with 5 μM RA were tested for neurotrophin-induced Trk tyrosine phosphorylation. Cells were stimulated with BDNF, NGF, NT-3, or NT-4/5 for 5 minutes and Trk proteins were immunoprecipitated with Trk antibody. The immunoprecipitates were analyzed by probing with phosphotyrosine antibody (panel 2).

protein (Figure 7B, panel 1). The expression of TrkA and TrkC was not induced by RA treatment. To determine whether RA induced the expression of functional TrkB receptors, we treated SY5Y cells with BDNF for 5 minutes followed by immunoprecipitation with an antibody that recognizes all known Trk family receptors. We then assayed the immunoprecipitated Trk proteins for ligand-induced tyrosine phosphorylation by probing with phosphotyrosine antibody. BDNF stimulated the tyrosine phosphorylation of TrkB in cells treated with RA for 2 days, with maximal responses at 8 days of RA treatment (Figure 7A, panel 2). Nanomolar concentrations of RA induced the expression of truncated $p95^{TrkB}$ and full-length $p145^{TrkB}$ (Figure 7B, panel 1), with a half-maximal dose of approximately 10 nM RA, which is similar to concentrations of RA found in responsive tissues. NT-3 and NT-4/5 also stimulated the tyrosine phosphorylation of TrkB in RA-treated SY5Y cells (Figure 7B, panel 2). Small amounts of Trk (TrkA) tyrosine phosphorylation were detected in SY5Y cells treated with both RA and NGF (Figure 7B, panel 2), consistent with the low amounts of Trk expression in these cells. These data show that RA induces the expression of functional TrkB receptors and suggest that RA induction of these receptors results in BDNF-mediated neuronal differentiation of SY5Y cells.

RA Induces the Expression of Truncated TrkB Receptors in NB Cells That Fail to Differentiate in Response to RA. In many human NB cell lines, RA treatment alone is sufficient to induce neuronal differentiation. We found that all of the cell lines that differentiate in response to RA constitutively expressed the TrkB ligand BDNF. Thus, in these cell lines, differentiation occurred in the absence of added BDNF. Several other NB cell lines failed to differentiate in response to either RA or RA together with BDNF. In one of these cell lines, 15N, which expressed BDNF, we found that RA induced the expression of a truncated, catalytically inactive form of TrkB. Unlike full-length TrkB, truncated TrkB lacks a tyrosine kinase domain and is not capable of transmitting neurotrophin signals. However, we were able to restore the ability of 15N cells to respond to RA by introducing a full-length TrkB receptor into these cells by transfection (Figure 8). These studies indicate that expression of full-length TrkB in 15N cells that express BDNF is sufficient to induce neuronal differentiation.

These results indicate that RA induces the neuronal differentiation of NB cells by modulating the expression of neurotrophin receptors. The effects of RA and BDNF on NB cells are shown in Figure 9. The induction of TrkB by RA and the acquisition of BDNF responsiveness suggest a fundamental role for RA in the development of nervous system cells.

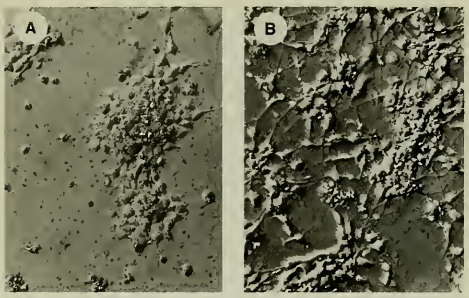


FIGURE 8. Morphology of TrkB-transfected 15N NB cells. 15N NB cells were transfected with control plasmid DNA (A) or plasmid encoding the full-length TrkB cDNA (B).

During embryonic development, RA may induce the expression of TrkB in neurons, thereby regulating the responsiveness to TrkB ligands. In addition, local production of RA may maintain TrkB levels in certain neurons. Thus, the response of neurons to neurotrophins may be regulated not only at the level of target-derived neurotrophin production but also at the level of receptor expression.

Therapeutic Potential of RA and BDNF. Retinoids are currently being used therapeutically to treat several forms of cancer. To more effectively use agents that induce differentiation as potential therapies for human cancers, it is necessary to understand the basic mechanisms controlling growth arrest and differentiation in tumor cells. Our studies describe a mechanism whereby the differentiation and growth arrest of NB tumor cells are induced by a combination of RA and exogenous or endogenous factors that stimulate the activity of TrkB. As most NB cell lines are derived from patients with advanced-stage disease and with a poor clinical prognosis, the results here support the integration of agents such as RA that induce high-level TrkB expression in the treatment of NB tumors. In the future, we will assess the therapeutic potential of RA and neurotrophins in the treatment of human cancer and evaluate the use of these agents to promote the growth arrest of other human brain tumor cells. We hope that the terminal differentiation of NB cells by RA and BDNF in combination with chemotherapy will provide an effective means of treatment for neuroblastoma.

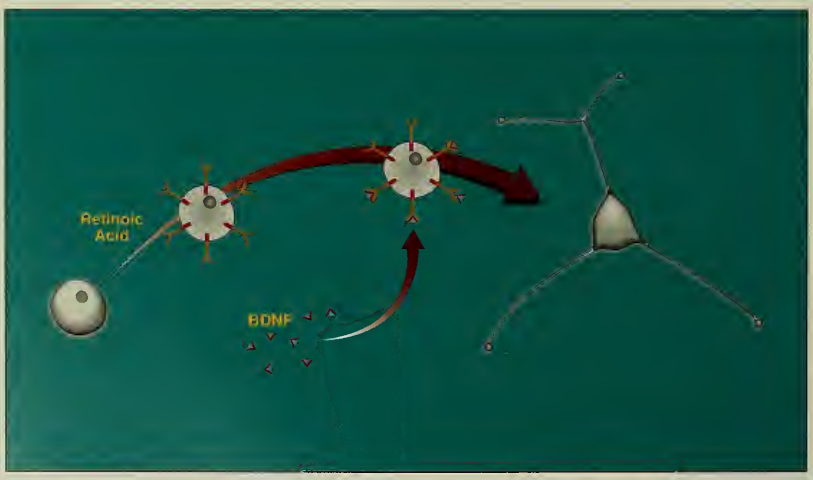
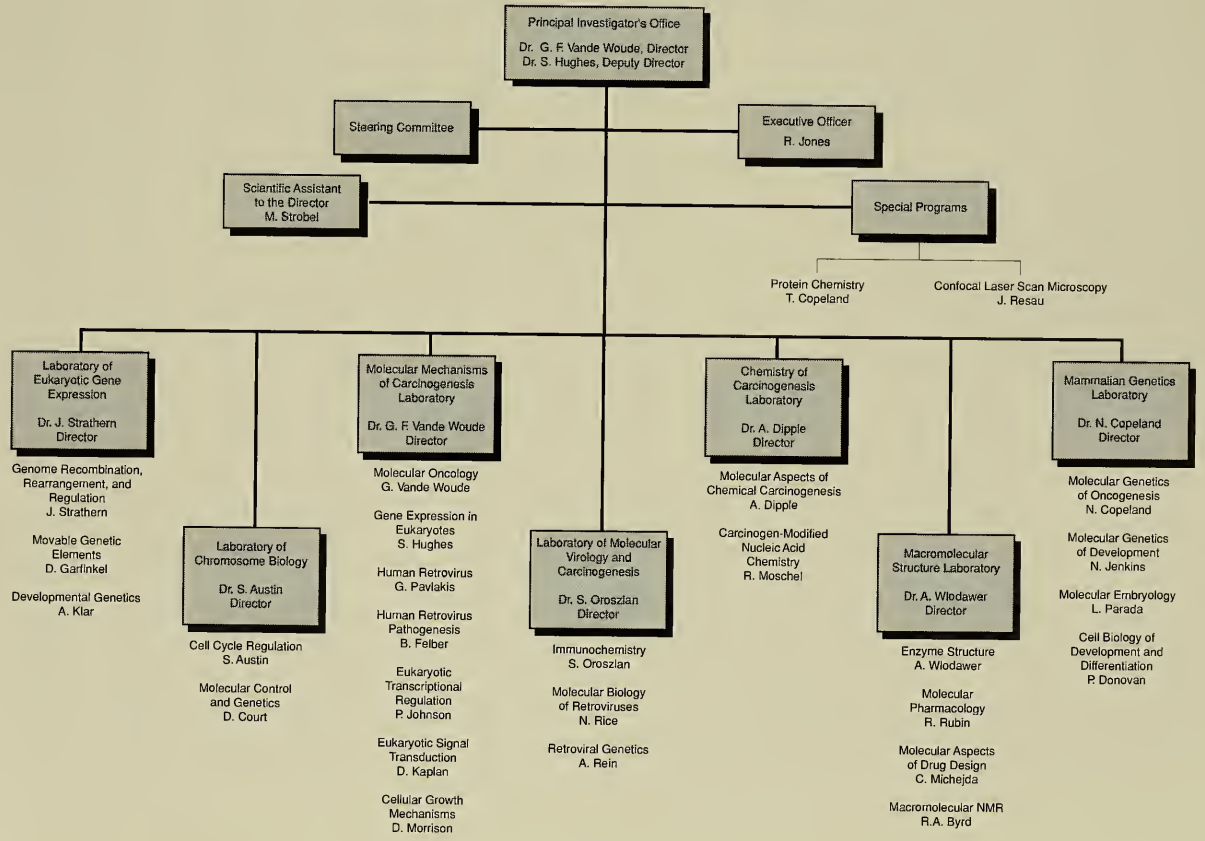


FIGURE 9. Mechanism of neuronal differentiation of NB cells. Human NB is a pediatric tumor that in later stages is malignant and nonresponsive to chemotherapy. We show that cells derived from NB tumors may be induced to stop growing and to differentiate into neuronal cells by treatment with RA and BDNF. RA induces the expression of the TrkB receptor tyrosine kinase. Treatment of the TrkB-expressing cells with the TrkB ligand BDNF stimulates the activity of TrkB and induces neuronal differentiation.

Staff Listings

Publications

NATIONAL CANCER INSTITUTE
FREDERICK CANCER RESEARCH AND DEVELOPMENT CENTER
ABL – BASIC RESEARCH PROGRAM



Scientific and Technical Staff

Ann Abeles, Scientist, Laboratory of Chromosome Biology. Education: University of Minnesota, 1963, BA (chemistry). University of Maryland, 1969, PhD (biophysics)

Jerry Alexandratos, Senior Research Technician, Macromolecular Structure Laboratory. Education: State University of New York at Stony Brook, 1986, BS (biochemistry). State University of New York at Buffalo, 1989, MA (biochemistry)

Stuart Austin, Director, Laboratory of Chromosome Biology, and Head, Cell Cycle Regulation Section. Education: University of London, 1968, BSc (microbiology). University of Edinburgh, 1971, PhD (microbiology)

Teresa Baker, Research Technician, Laboratory of Chromosome Biology. Education: Hood College, 1985, BA (biology)

Donna Baldisseri, Scientist Associate, Macromolecular Structure Laboratory. Education: Boston University, 1981, BA (chemistry). Wesleyan University, 1988, PhD (chemistry)

Deborah Barnhart, Research Technician, Mammalian Genetics Laboratory. Education: Eastern Nazarene College, 1990, BS (biology)

Mary Barnstead, Research Technician, Mammalian Genetics Laboratory. Education: Virginia Polytechnic Institute and State University, 1990, BS (biology)

Robin Bhaerman, Senior Research Technician, Molecular Mechanisms of Carcinogenesis Laboratory. Education: Duke University, 1989, BA (philosophy)

Michael Bonaduce, Senior Research Technician, Laboratory of Eukaryotic Gene Expression. Education: Walsh College, 1985, BS (biology). University of Richmond, 1988, MS (genetics)

Therese Brendler, Scientist Associate, Laboratory of Chromosome Biology. Education: University of Texas at Austin, 1974, BS (chemistry). Washington University of St. Louis, 1982, PhD (molecular biology)

Carlton Briggs, Senior Research Technician, Laboratory of Molecular Virology and Carcinogenesis. Education: Howard University, 1971, BS (microbiology); 1991, MS (microbiology)

Stanley Brown, Scientist Associate, Laboratory of Chromosome Biology. Education: University of Massachusetts, 1973, BS (microbiology). Indiana University, 1975, MA (microbiology). Harvard University, 1982, PhD (microbiology, molecular genetics)

Thomas Burkett, Doctoral Student, Laboratory of Eukaryotic Gene Expression. Education: Western Washington University, 1984, BS (chemistry, biological sciences). University of Maryland at Baltimore County, 1988, MS (applied molecular biology)

R. Andrew Byrd, PhD, Head, Macromolecular NMR Section, Macromolecular Structure Laboratory. Education: Furman University, 1973, BS (chemistry). University of South Carolina, 1977, PhD (physical chemistry)

Carrie Cantwell, Research Technician, Molecular Mechanisms of Carcinogenesis Laboratory Education: Michigan State University, 1988, BS (biochemistry)

Melonie Cassell, Senior Research Technician, Laboratory of Molecular Virology and Carcinogenesis

Brian Cho, Research Technician, Mammalian Genetics Laboratory. Education: Virginia Polytechnic Institute and State University, 1987, BS (biology, microbiology). Hood College, 1993, MS (biomedical sciences)

W. Marek Cholody, Scientist Associate, Macromolecular Structure Laboratory. Education: Technical University of Gdansk, Poland, 1977, MSc Eng (chemistry); 1986, PhD (chemistry)

Linda Cleveland, Senior Research Technician, Mammalian Genetics Laboratory. Education: Hood College, 1982, BS (biology)

Debra Compton, Senior Research Technician, Mammalian Genetics Laboratory. Education: Norfolk State University, 1989, BS (environmental toxicology)

Pantelis Constantoulakis, Scientist Associate, Molecular Mechanisms of Carcinogenesis Laboratory. Education: University of Athens, 1983, BS (biology); 1985, MS (marine biology); 1987, MD (medicine). University of Washington, 1990, PhD (medical genetics)

Neal Copeland, Director, Mammalian Genetics Laboratory, and Head, Molecular Genetics of Oncogenesis Section. Education: University of Utah, 1971, BS (biology); 1976, PhD (biochemistry)

Terry Copeland, Head, Special Program in Protein Chemistry. Education: University of Texas, 1961, BS (chemistry)

Nina Costantino, Senior Research Technician, Laboratory of Chromosome Biology. Education: Hood College, 1979, BS (medical technology)

Donald Court, Head, Molecular Control and Genetics Section, Laboratory of Chromosome Biology. Education: Colgate University, 1965, BA (physical sciences). Rochester University, 1971, PhD (biology)

Christian Cunningham, Senior Research Technician, Molecular Mechanisms of Carcinogenesis Laboratory. Education: Mount Saint Mary's College, 1982, BS (biology)

Avis Danishefsky, Scientist Associate, Macromolecular Structure Laboratory. Education: Barnard College, 1979, AB (biochemistry). City University of New York, 1987, PhD (biochemistry)

Santanu DasGupta, Scientist Associate, Laboratory of Chromosome Biology. Education: Calcutta University, 1965, BSc (physics, chemistry, mathematics); 1968, MSc (physics, biophysics); 1973, PhD (biophysics)

Michael Davis, Scientist Associate, Laboratory of Chromosome Biology. Education: University of California at Berkeley, 1978, AB (bacteriology). Harvard University, 1986, PhD (biochemistry, molecular biology)

Leslie Derr, Scientist Associate, Laboratory of Eukaryotic Gene Expression. Education: Drew University, 1979, BA (psychobiology). Duke University, 1989, PhD (genetics)

Anthony Dipple, Director, Chemistry of Carcinogenesis Laboratory, and Head, Molecular Aspects of Chemical Carcinogenesis Section. Education: University of Birmingham, 1961, BSc (chemistry); 1964, PhD (biological chemistry)

Peter Donovan, Leader, Cell Biology of Development and Differentiation Group, Mammalian Genetics Laboratory. Education: University College, London, 1978, BSc (zoology, comparative anatomy); 1988, PhD (cell biology)

Paul Dowling, Jr., Scientific Programmer, Macromolecular Structure Laboratory. Education: Embry-Riddle Aeronautical University, 1991, BS (computer science)

Linda Dunn, Senior Research Technician, Laboratory of Molecular Virology and Carcinogenesis. Education: Pennsylvania State University, 1982, BS (microbiology)

Bryn Eagleson, Research Technician, Mammalian Genetics Laboratory. Education: Frederick Community College, 1989, AA (general studies)

Mary Ernst, Senior Research Technician, Laboratory of Molecular Virology and Carcinogenesis. Education: Hood College, 1968, BA (biology); 1988, MS (biomedical science)

Brenda Faiola, Research Technician, Laboratory of Eukaryotic Gene Expression. Education: Virginia Polytechnic Institute and State University, 1991, BS (biochemistry, chemistry)

Donna Faletto, Scientist Associate, Molecular Mechanisms of Carcinogenesis Laboratory. Education: Bucknell University, 1982, BS (chemistry). University of Rochester, 1986, MS (toxicology); 1988, PhD (toxicology)

David Farnsworth, Senior Research Technician, Macromolecular Structure Laboratory. Education: Shepherd College, 1986, BS (chemistry)

Mark Federspiel, Scientist Associate, Molecular Mechanisms of Carcinogenesis Laboratory. Education: University of Dayton, 1980, BS (biology). Michigan State University, 1987, PhD (biochemistry)

Barbara Felber, Leader, Human Retrovirus Pathogenesis Group, Molecular Mechanisms of Carcinogenesis Laboratory. Education: Kantonsschule Luzern, 1971, Matura (Typus C). University of Bern, 1977, MS (biology); 1980, PhD (molecular biology)

Ya-Xiong Feng, Scientist Associate, Laboratory of Molecular Virology and Carcinogenesis. Education: Shanghai Institute of Biochemistry, 1978, PhD (biochemistry)

Andrea Ferris, Senior Research Technician, Molecular Mechanisms of Carcinogenesis Laboratory. Education: Virginia Polytechnic Institute and State University, 1979, BS (biology)

Cara Fiore, Research Technician, Molecular Mechanisms of Carcinogenesis Laboratory. Education: Tulane University, 1989, BS/BA (biology/anthropology)

Colin Fletcher, Scientist Associate, Mammalian Genetics Laboratory. Education: Dartmouth College, 1980, AB (biochemistry). Rockefeller University, 1988, PhD (biochemistry, molecular biology)

David Garfinkel, Head, Movable Genetic Elements Section, Laboratory of Eukaryotic Gene Expression. Education: University of Wisconsin, 1975, BS (bacteriology). University of Washington, 1977, MS (microbiology); 1981, PhD (microbiology)

Debra Gilbert, Senior Research Technician, Mammalian Genetics Laboratory. Education: Western Maryland College, 1981, BS (biology); Hood College, 1987, MS (biomedical science)

Iris Givol, Senior Research Technician, Molecular Mechanisms of Carcinogenesis Laboratory. Education: Hebrew University, 1977, BSc (chemistry)

Ganesan Gopalan, Scientist Associate, Mammalian Genetics Laboratory. Education: Madurai Kamaraj University, 1979, MSc (biochemistry); 1986, PhD (biochemistry)

Jacqueline Gordon, Senior Research Technician, Laboratory of Molecular Virology and Carcinogenesis. Education: Frostburg State University, 1984, BS (biology). Kansas State University, 1990, MS (microbiology)

Deanna Gotte, Senior Research Technician, Laboratory of Molecular Virology and Carcinogenesis. Education: Northern Arizona University, 1984, BS (biology); 1991, MS (applied microbiology)

Alla Gustchina, Scientist Associate, Macromolecular Structure Laboratory. Education: University of Moscow, 1971, BS (chemistry). Institute of Molecular Biology (Moscow), 1984, PhD (molecular biology)

Tadeusz Guszczynski, Scientist Associate, Laboratory of Molecular Virology and Carcinogenesis. Education: Medical University of Wroclaw, Poland, 1980, MS (pharmacy); 1989, PhD (biochemistry)

Jeanette Ewald Harrison, Senior Research Technician, Molecular Mechanisms of Carcinogenesis Laboratory. Education: Shepherd College, 1986, BS (biology)

Demetria Harvin, Research Technician, Laboratory of Molecular Virology and Carcinogenesis. Education: Hampton Institute, 1986, BA (biology)

Lidia Hernandez, Scientist Associate, Macromolecular Structure Laboratory. Education: Millersville State College, 1975, BA (biology); 1979, MS (biology)

Catherine Hixson, Senior Research Technician, Laboratory of Molecular Virology and Carcinogenesis. Education: Frostburg State University, 1973, BS (biology)

Stephen Hughes, Deputy Director, Basic Research Program, and Head, Gene Expression in Eukaryotes Section, Molecular Mechanisms of Carcinogenesis Laboratory. Education: Washington University, 1969, BA (biology). Harvard University, 1975, PhD (biochemistry)

Nancy Jenkins, Head, Molecular Genetics of Development Section, Mammalian Genetics Laboratory. Education: Sweet Briar College, 1972, BA (chemistry). Indiana University, 1974, MS (biochemistry); 1977, PhD (molecular and cellular biology)

Peter Johnson, Leader, Eukaryotic Transcriptional Regulation Group, Molecular Mechanisms of Carcinogenesis Laboratory. Education: University of Colorado, 1975, BA (biology). University of California at San Diego, 1983, PhD (biology)

David Kaplan, Leader, Eukaryotic Signal Transduction Group, Molecular Mechanisms of Carcinogenesis Laboratory. Education: Clark University, 1978, BA (biology). Harvard University, 1987, PhD (cell and developmental biology)

Young Kim, Senior Research Technician, Laboratory of Molecular Virology and Carcinogenesis. Education: Kyung Hee University, 1971, BS (chemistry)

Amar J.S. Klar, Head, Developmental Genetics Section, Laboratory of Eukaryotic Gene Expression. Education: Punjab Agriculture University, 1967, BS (biochemistry); 1969, MS (microbiology). University of Wisconsin at Madison, 1975, PhD (bacteriology, molecular biology)

Marilyn Kroeger Koepke, Scientist II, Macromolecular Structure Laboratory. Education: St. Olaf College, 1973, BA (chemistry). University of Nebraska, 1976, MS (chemistry); 1978, PhD (chemistry)

Leslie Lock, Scientist Associate, Mammalian Genetics Laboratory. Education: San Diego State University, 1975, BA (biology); 1977, MS (genetics). University of California at San Francisco, 1985, PhD (anatomy)

Eva Majerova, Research Technician, Special Program in Protein Chemistry

Karen Mathes, Research Technician, Molecular Mechanisms of Carcinogenesis Laboratory. Education: Gettysburg College, 1992, BS (biology)

Marie Mazzulla, Senior Research Technician, Mammalian Genetics Laboratory. Education: York College/CUNY, 1981, BS (chemistry)

Dwayne McCully, Scientific Programmer, Macromolecular Structure Laboratory. Education: University of Maryland, 1989, BS (computer science)

Mary Jane McWilliams, Senior Research Technician, Molecular Mechanisms of Carcinogenesis Laboratory. Education: Mercyhurst College, 1975, BA (biology, chemistry)

Carolyn McGill, Senior Research Technician, Laboratory of Eukaryotic Gene Expression. Education: Lycoming College, 1969, BA (biology)

Christopher Michejda, Head, Molecular Aspects of Drug Design Section, Macromolecular Structure Laboratory. Education: University of Illinois, 1959, BS (chemistry). University of Rochester, 1963, PhD (physical organic chemistry)

Linda Miller, Senior Research Technician, Molecular Mechanisms of Carcinogenesis Laboratory

Maria Miller, Scientist, Macromolecular Structure Laboratory. Education: University of Warsaw, 1971, MS (physics); 1979, PhD (biophysics)

Jane Mirro, Senior Research Technician, Laboratory of Molecular Virology and Carcinogenesis. Education: Pennsylvania State University, 1966, BS (medical technology). University of Maryland, 1974, MS (microbiology)

Karen Moore, Scientist Associate, Mammalian Genetics Laboratory. Education: Leicester University, 1977, BSc (biological science); 1981, PhD (genetics)

Sharon Moore, Scientist Associate, Laboratory of Eukaryotic Gene Expression. Education: Morris Harvey College, 1972, BS (biology). Western Kentucky University, 1974, MS (biology). University of Louisville, 1982, PhD (biology)

Deborah Morrison, Leader, Cellular Growth Mechanisms Group, Molecular Mechanisms of Carcinogenesis Laboratory. Education: David Lipscomb College, 1978, BS (biology, chemistry). Vanderbilt University, 1985, PhD (microbiology)

Robert Moschel, Head, Carcinogen-Modified Nucleic Acid Chemistry Section, Chemistry of Carcinogenesis Laboratory. Education: Ohio State University, 1968, BS (biochemistry); 1973, PhD (biochemistry)

Stephen Oroszlan, Director, Laboratory of Molecular Virology and Carcinogenesis, and Head, Immunochemistry Section. Education: Technical University, Budapest, 1950, MS (chemical engineering). Georgetown University, 1950, PhD (pharmacology)

Marianne Oskarsson, Scientist II, Molecular Mechanisms of Carcinogenesis Laboratory. Education: College of Home Economics, Stockholm, Sweden, 1959, Diploma. University of Maryland, 1968, MS (food science)

Norene O'Sullivan, Research Technician, Mammalian Genetics Laboratory. Education: Northern Virginia Community College, 1989, AS (engineering). University of Maryland at College Park, 1991, BS (biology, genetics)

John Page, Research Technician, Chemistry of Carcinogenesis Laboratory. Education: University of Maryland, 1987, BS (biological sciences, genetics)

Mary Ellen Palko, Research Technician, Mammalian Genetics Laboratory. Education: University of Michigan, 1988, BS (microbiology)

Luis F. Parada, Head, Molecular Embryology Section, Mammalian Genetics Laboratory. Education: University of Wisconsin at Madison, 1979, BS (molecular biology). Massachusetts Institute of Technology, 1985, PhD (biology)

Gary Pauly, Scientist Associate, Chemistry of Carcinogenesis Laboratory. Education: University of Cincinnati, 1982, BA (psychology); 1987, PhD (biochemistry)

George Pavlakis, Head, Human Retrovirus Section, Molecular Mechanisms of Carcinogenesis Laboratory. Education: National University of Athens Medical School, 1976, MD. Syracuse University, 1980, PhD (biology)

Alexander Pavlovsky, Scientist Associate, Macromolecular Structure Laboratory. Education: Moscow University, 1971, MS (geology); 1979, PhD (molecular biology)

Stuart Rabin, Senior Research Technician, Molecular Mechanisms of Carcinogenesis Laboratory. Education: Fairleigh Dickinson University, 1986, BS (biology). University of Delaware, 1991, MS (animal science)

J.K. Mohana Rao, Scientist Associate, Macromolecular Structure Laboratory. Education: Sri Vankateswara University, 1960, BSc (physics, chemistry, mathematics); 1962, MSc (physics). Indian Institute of Science, 1971, PhD (X-ray crystallography)

Lucretia Reaves, Senior Research Technician, Laboratory of Chromosome Biology. Education: Shepherd College, 1963, BS (biology)

Susan Reid, Research Technician, Mammalian Genetics Laboratory. Education: North Carolina State University, 1986, BS (animal science)

Alan Rein, Head, Retroviral Genetics Section, Laboratory of Molecular Virology and Carcinogenesis. Education: Reed College, 1961, BA (biology). University of California at Berkeley, 1967, PhD (molecular biology)

James Resau, Head, Special Progam in Confocal Laser Scan Microscopy. Education: Western Maryland College, 1968, BA (history). University of Maryland, 1977, MS (pathology); 1985, PhD (pathology)

James Resnick, Scientist Associate, Mammalian Genetics Laboratory. Education: Colgate University, 1976, BA (biology). University of Pittsburgh, 1982, PhD (biological science)

Nancy Rice, Head, Molecular Biology of Retroviruses Section, Laboratory of Molecular Virology and Carcinogenesis. Education: Stanford University, 1961, BA (biology). Harvard University, 1963, MA (biology); 1969, PhD (biology)

April Rollins, Research Technician, Macromolecular Structure Laboratory. Education: Hope College, 1988, BS (biology)

Sing Rong, Scientist Associate, Molecular Mechanisms of Carcinogenesis Laboratory. Education: Shandong University, 1982, BS (microbiology). Tufts University School of Medicine, 1989, PhD (molecular biology)

Ronald Rubin, Leader, Molecular Pharmacology Group, Macromolecular Structure Laboratory. Education: University of Wisconsin at Madison, 1969, BS (chemistry); 1974, PhD (molecular biology)

Shen Rulong, Scientist Associate, Molecular Mechanisms of Carcinogenesis Laboratory. Education: Zhangjiang Medical College, 1981, BS (medicine). Nanjing Medical College, 1986, MD (tumor pathology)

Cleopas Samudzi, Scientist Associate, Macromolecular Structure Laboratory. Education: Dakota Wesleyan University, 1980, BA (chemistry, biology). University of Pittsburgh, 1983, MS (biochemistry, biophysics); 1989, PhD (biological sciences)

Bangalore Sathyanarayana, Scientific Programmer, Macromolecular Structure Laboratory. Education: Mysone University, 1964, BSc (physics, mathematics); 1966, MSc (physics). Madras University, 1972, PhD (physics)

Nicholas Schulz, Scientist Associate, Molecular Mechanisms of Carcinogenesis Laboratory. Education: University of Pittsburgh, 1978, BS (chemistry). University of Pittsburgh School of Medicine, 1982, MD

Lisa Secrest, Research Technician, Mammalian Genetics Laboratory

Brenda Shafer, Research Technician, Laboratory of Eukaryotic Gene Expression

Jeffrey Strathern, Director, Laboratory of Eukaryotic Gene Expression, and Head, Genome Recombination, Rearrangement, and Regulation Section. Education: University of California at San Diego, 1970, BA (biology). University of Oregon, 1977, PhD (biology)

Pramod Sutrave, Scientist Associate, Molecular Mechanisms of Carcinogenesis Laboratory. Education: Osmania University, 1974, BS (botany). Jawaharlal Nehru University, 1980, PhD (microbiology)

Amy Swain, Scientist Associate, Macromolecular Structure Laboratory. Education: Frostburg State University, 1983, BS (biology). University of South Carolina, 1988, PhD (chemistry)

Deborah Swing, Scientist, Mammalian Genetics Laboratory. Education: University of Cincinnati, 1981, BS (biology)

Nadezhda Tarasova, Scientist Associate, Macromolecular Structure Laboratory. Education: Moscow State University, Russia, 1980, PhD (chemistry)

James Turpin, Scientist Associate, Molecular Mechanisms of Carcinogenesis Laboratory. Education: University of Evansville, 1980, BS (biology, chemistry). University of Texas Graduate School of Biomedical Sciences, 1984, MS (biological sciences); 1988, PhD (biological sciences, immunology)

Tami Unangst-Steyskal, Research Technician, Laboratory of Molecular Virology and Carcinogenesis. Education: Hood College, 1991, BS (chemistry)

George Vande Woude, Principal Investigator, Basic Research Program, Director, Molecular Mechanisms of Carcinogenesis Laboratory, and Head, Molecular Oncology Section. Education: Hofstra University, 1959, BA (biochemistry). Rutgers University, 1962, MS (biochemistry); 1964, PhD (biochemistry)

Deborah Wancio, Research Technician, Molecular Mechanisms of Carcinogenesis Laboratory. Education: Gettysburg College, 1992, BS (biology)

Keith Weinstock, Scientist Associate, Laboratory of Eukaryotic Gene Expression. Education: University of California at Davis, 1977, BS (botany); 1979, MS (botany). University of California at Berkeley, 1985, PhD (organic chemistry)

Patricia Winberg, Research Technician, Mammalian Genetics Laboratory. Education: Johns Hopkins University, 1990, BS (biology)

Alexander Wlodawer, Director, Macromolecular Structure Laboratory, and Head, Enzyme Structure Section. Education: University of Warsaw, 1968, MS (physics). University of California at Los Angeles, 1974, PhD (molecular biology)

Jian-Cheng Xuan, Scientist Associate, Macromolecular Structure Laboratory. Education: Shanghai Institute of Biochemistry, 1987, PhD (biochemistry)

Nelson Yew, Scientist Associate, Molecular Mechanisms of Carcinogenesis Laboratory. Education: University of Michigan, 1982, BS (cell and molecular biology). Northwestern University, 1988, PhD (biochemistry, molecular biology, and cell biology)

Brenda Youngren-Grimes, Senior Research Technician, Laboratory of Chromosome Biology. Education: Mankato State University, 1973, BS (medical technology). University of Wisconsin at La Crosse, 1981, MS (biology, microbiology)

Renping Zhou, Scientist Associate, Molecular Mechanisms of Carcinogenesis Laboratory. Education: Nanjing Teacher's School, 1982, BS (biology). University of California at Berkeley, 1988, PhD (molecular biology)

Administrative and Support Staff

Anne Arthur, Senior Technical Editor/Writer, Administrative Office

Alisa Blackledge, Executive Secretary, Molecular Mechanisms of Carcinogenesis Laboratory

Linda Brubaker, Administrative Assistant II, Mammalian Genetics Laboratory

Sondra Cavanaugh, Senior Accountant, Administrative Office

Patricia Chlop, Secretary, Chemistry of Carcinogenesis Laboratory

Rosanne Cuttitta, Personnel Administrator, Administrative Office

Kelly Eichelberger, Clerk Typist, Director's Office

Cynthia Fitzpatrick, Senior Secretary, Mammalian Genetics Laboratory

Roberta Greenhouse, Administrative Specialist I, Molecular Mechanisms of Carcinogenesis Laboratory

Catherine Hahn, Senior Secretary, Mammalian Genetics Laboratory

Robin Handley, Senior Secretary, Molecular Mechanisms of Carcinogenesis Laboratory

Deborah Higdon, Personnel Associate, Administrative Office

Joan Hopkins, Executive Secretary, Laboratory of Eukaryotic Gene Expression

Stephen Hughes, PhD, Deputy Director, Basic Research Program

David Erin Jenkins, Administrative Specialist II, Administrative Office

Roberta Jones, Executive Officer, Basic Research Program

Laraine Kasunic, Editorial Assistant, Administrative Office

Kathleen Kasten, Senior Secretary, Macromolecular Structure Laboratory

Susan Kelly, Executive Secretary, Macromolecular Structure Laboratory

Sylvia Lucas, Secretary, Administrative Office

Hilda Marusiodis, Administrative Assistant I, Molecular Mechanisms of Carcinogenesis Laboratory

JoAnne Muldoon, Executive Secretary, Macromolecular Structure Laboratory

Ruby Poulin, Senior Clerk Typist, Molecular Mechanisms of Carcinogenesis Laboratory

Julie Ratliff, Administrative Assistant I, Laboratory of Chromosome Biology

Michelle Reed, Executive Secretary, Molecular Mechanisms of Carcinogenesis Laboratory

Cheri Rhoderick, Administrative Assistant I, Laboratory of Molecular Virology and Carcinogenesis

Erika Ridgway, Administrative Specialist I, Macromolecular Structure Laboratory

Carol Shawver, Executive Secretary, Laboratory of Molecular Virology and Carcinogenesis

Marjorie Strobel, PhD, Scientific Assistant to the Director, Basic Research Program

Lori Summers, Administrative Specialist to the Director, Basic Research Program

George Vande Woude, PhD, Principal Investigator and Director, Basic Research Program

Carol Whipp, Administrative Assistant I, Chemistry of Carcinogenesis Laboratory

Diane Whiten, Administrative Assistant II, Administrative Office

Madeline Wilson, Executive Secretary, Molecular Mechanisms of Carcinogenesis Laboratory

Visiting Scientist Program

Researchers from the United States and abroad participate in the Basic Research Program through the Visiting Scientist Program. Scientists from other institutions are hosted for periods of several days to a year or more. These exchanges are funded from various sources. Visiting Scientists are usually supported in full or in part by the Basic Research Program. Guest Workers are funded by outside sources of support such as grants, fellowships, and sabbaticals. The following researchers visited the Basic Research Program during the reporting period.

C. Anita Bigger, PhD, Guest Worker from Microbiological Associates, Rockville, MD. Research on mutational specificity—host investigator, Anthony Dipple

Jeffrey Blumenstein, PhD, Guest Worker from the Food and Drug Administration, Rockville, MD. Research on drug design—host investigator, Christopher Michejda

Thomas Cameron, Guest Worker from Penn State University, Abington, PA. Research on computational studies of the structures of bistriazenes—host investigator, Christopher Michejda

Su-Min Chen, MD, Visiting Scientist from the Xian Institute of Molecular Biology, Xian, People's Republic of China. Research on protein-protein interaction, Era protein, and artificial gene selection—host investigator, Donald Court

Joseph Ciotola, Guest Worker from Western Maryland College, Westminster, MD. Research on studies of triazene metabolism by rat hepatocytes—host investigator, Christopher Michejda

Amikam Cohen, PhD, Visiting Scientist from The Hebrew University, Jerusalem, Israel. Research on mating-type switching in yeast, break-induced recombination, control of mating-type switching—host investigator, Amar J.S. Klar

Daniel Court, Guest Worker. Research on PCR analysis, plasmid construction, and other recombinant DNA techniques—host investigator, Stuart Austin

Eva Csak, MD, Guest Worker from Koranyi Korhaz, Kozponti Labratorium, Budapest, Hungary. Instruction in immunochemical methods for identification of HIV proteins—host investigator, Stephen Oroszlan

Asis Das, PhD, Guest Worker from the University of Connecticut Health Center, Farmington, CT. Research on the regulation of transcription termination in bacteriophage lambda—host investigator, Donald Court

Alun Davies, PhD, Guest Worker from St. Georges Hospital Medical School, London, England. Research on N*myc* mutant embryos—host investigator, Luis F. Parada

Susanna Dolci, PhD, Guest Worker from the University of Rome, Italy. Research on the growth factor receptors involved in the development of the mouse germ line—host investigator, Peter Donovan

George Gause, PhD, Guest Worker from the Institute of Developmental Biology, Russian Academy of Sciences, Moscow, Russia. Research on selected cellular and viral target proteins for the development of specific binding and binding-inhibition assays—host investigator, George Vande Woude

Raymond Gilden, PhD, Senior Guest Investigator from Program Resources, Inc., NCI-FCRDC. Research on retroviral pathogenesis—host investigator, Nancy Rice

Emilia Hadziyannis, MD, Visiting Scientist from the University of Athens, Athens, Greece. Research on the molecular biology of human retroviruses and AIDS—host investigator, George Pavlakis

Amnon Hizi, PhD, Visiting Scientist from Tel Aviv University, Tel Aviv, Israel. Research on the reverse transcriptases of HIV-1 and HIV-2—host investigator, Stephen Hughes

Mariusz Jaskolski, PhD, Visiting Scientist from A. Mickiewicz University, Poznan, Poland. Research on crystallographic studies of protein structure—host investigator, Alexander Wlodawer

Iafa Keydar, PhD, Guest Worker from Tel Aviv University, Israel. Research on the localization of the c-*met* product in breast tumors—host investigator, George Vande Woude

Uwe Klemm, PhD, Guest Worker from the Whitehead Institute of the Massachusetts Institute of Technology, Cambridge, MA. Research on primordial germ cell development—host investigator, Peter Donovan

Steven Koepke, PhD, Guest Worker from the Environmental Protection Agency, Arlington, VA. Research on drug metabolism—host investigator, Christopher Michejda

Britta Koppe, MD, Visiting Scientist from the Institute of Pathology, Aachen, Germany. Research on inhibitors of the protease of HIV-I—host investigator, Stephen Oroszlan

Ryoko Kuriyama, PhD, Guest Worker from the University of Minnesota, Minneapolis, MN. Research on phosphorylation of microtubule-associated proteins—host investigator, George Vande Woude

Sally Look, PhD, Guest Worker from the Food and Drug Administration, Rockville, MD. Research on modeling of drug-DNA interactions—host investigator, Christopher Michejda

Molly Mastrangelo, PhD, Guest Worker from Allegany Community College, Cumberland, MD. Research on silencing of the *HML*α locus—host investigator, Jeffrey Strathern

Vladimir Morozov, PhD, Visiting Scientist from the Cancer Research Center of Russian Federation, Moscow, Russia. Research on protease inhibitors in HTLV-1– and HIV–infected cell cultures—host investigator, Stephen Oroszlan

Karoly Nagy, PhD, Visiting Scientist from the Institute of Isotopes, Hungarian Academy of Sciences, Budapest, Hungary. Research on virus-cell interactions and early events of human retroviral infections—host investigator, Stephen Oroszlan

Takuro Nakamura, PhD, Guest Worker from the University of Tokyo, Faculty of Medicine, Tokyo, Japan. Research on carcinogenesis and embryogenesis—host investigator, Neal Copeland

Markus W.F. Neumann, PhD, Visiting Scientist from the GSF Research Center for Environment and Health, Neuherberg, Germany. Research on the expression of HIV-1 in human brain cells—host investigator, George Pavlakis

Irene Potouridou, MD, Visiting Scientist from A. Sygros Hospital, Athens, Greece. Research on AIDS—host investigator, George Pavlakis

Carol Rouzer, PhD, Guest Worker from Western Maryland College, Westminster, MD. Research on alkyltriazenes as potential cancer therapeutic agents—host investigator, Christopher Michejda

Ann Schmiedekamp, PhD, Visiting Scientist from Pennsylvania State University, Abington, PA. Research on modeling of new chemotherapeutic drugs—host investigator, Christopher Michejda

Karen Schueler, Guest Worker. Research on the preparation and purification of RNA from tissue and plasmid DNA used in RNase protection and in situ hybridization assays—host investigator, Stephen Hughes

Mark Schwartz, Guest Worker from Pennsylvania State University, Abington, PA. Research on computational studies on the structures of bistriazenes—host investigator, Christopher Michejda

Shraga Segal, PhD, Visiting Scientist from the Ben-Gurion University of the Negev, Beer-Sheva, Israel. Research on immunological and molecular mechanisms of cancer metastases—host investigator, George Vande Woude

Kyoichi Shibuya, PhD, Guest Worker from the Nippon Kayaku Company, Ltd., Tokyo, Japan. Research on signal transduction—host investigator, David Kaplan

Linda Siracusa, PhD, Guest Worker from the Fels Institute, Philadelphia, PA. Research on in situ hybridization of ALL-1 expression during mouse embryogenic development—host investigator, Luis F. Parada

Richard Smith, PhD, Visiting Scientist from Western Maryland College, Westminster, MD. Research on dialkyltriazenes and derivatives as potential carcinogenic and cancer chemotherapeutic agents—host investigator, Christopher Michejda

Nechama Smorodinsky, PhD, Guest Worker from Tel Aviv University, Tel Aviv, Israel. Research on the procedure of generating mouse monoclonal antibodies against Met and Mos—host investigator, George Vande Woude

Kenji Suzuki, PhD, Guest Worker from CIBA-GEIGY Ltd., Takarazuka, Japan. Research on the structure–function analysis of Trk receptors—host investigator, Luis F. Parada

Lisa Taneyhill, Guest Worker from Western Maryland College, Westminster, MD. Research on the interaction of chemotherapeutic drugs with DNA—host investigator, Christopher Michejda

Yoshio Taniyama, PhD, Guest Worker from Biology Research Laboratories, Takeda Chemical Industries Ltd., Osaka, Japan. Research on the structure–function analysis of Trk receptors—host investigator, Luis F. Parada

Leanne Tesene, Guest Worker from National Veterinary Services Laboratory, Ames, Iowa. Research on retroviral vectors—host investigator, Mark Federspiel

Jozsef Tozser, PhD, Visiting Scientist from the Medical University of Debrecen, Debrecen, Hungary. Research on the substrate specificity of HIV, EIAV, and HTLV proteinases—host investigator, Stephen Oroszlan

Alfons Valera, PhD, Guest Worker from the University of Barcelona, Barcelona, Spain. Research to dissect the regulatory sequences of the TrkB receptor at the sites of action for the NGF-related neurotrophins—host investigator, Luis F. Parada

Agneta von Gegerfelt, PhD, Guest Worker from the Karolinska Institute, Stockholm, Sweden. Research on the molecular biology of HIV-1 regulatory elements—host investigator, George Pavlakis

Jian-guang Zhou, PhD, Visiting Scientist from the Molecular Genetics Center, Beijing, People's Republic of China. Research on the mechanism of transcription termination and antitermination by RNA polymerase during elongation—host investigator, Donald Court

Postdoctoral Fellowship Program

The Basic Research Program is committed to providing excellent postdoctoral training to recipients of an MD, PhD, or equivalent degree in the biological or biochemical sciences. The fellowship is funded by the laboratory to which the postdoctoral fellow is assigned, by Basic Research Program Fellowship funds (Director's funds), or by outside sources. Appointments are made for 1 to 3 years. Each postdoctoral fellow is assigned to a senior staff member who oversees the progress and direction of research. The following postdoctoral fellows worked in the Program laboratories during the reporting year.

Yoshifumi Adachi, Molecular Mechanisms of Carcinogenesis Laboratory. Education: Tohoku University, 1984, BS Agr (agricultural chemistry); 1986, MS Agr (agricultural chemistry). Kyoto University, 1988, D Med Sci (medical science)

Elena Afonina, Molecular Mechanisms of Carcinogenesis Laboratory. Education: Moscow State University, 1983, Diploma (chemistry); 1991, PhD (chemistry)

Rajiv Agarwal, Chemistry of Carcinogenesis Laboratory. Education: Rohilkriand University, 1983, MSc (chemistry). Aligarh Muslim University, 1989, PhD (chemistry). Queen's University of Belfast, 1992, PhD (chemistry)

Amanda Altieri, Macromolecular Structure Laboratory. Education: San Diego State University, 1986, BS (chemistry/biochemistry). University of North Carolina, 1993, PhD (chemistry)

Stephen Angeloni, Molecular Mechanisms of Carcinogenesis Laboratory. Education: Florida State University, 1980, BS (biology); 1984, MSc (molecular biology). Virginia Polytechnic Institute and State University, 1992, PhD (biochemistry/nutrition)

Karen Avraham, Mammalian Genetics Laboratory. Education: Washington University of St. Louis, 1984, BA (biology). Feinberg Graduate School, Weizmann Institute of Science, 1991, PhD (biology)

Mark E. Baer, Basic Research Program Postdoctoral Fellowship, Molecular Mechanisms of Carcinogenesis Laboratory. Education: Cornell University, 1986, BA (biochemistry). University of North Carolina at Chapel Hill, 1993, PhD (biochemistry)

Eugene Barsov, Molecular Mechanisms of Carcinogenesis Laboratory. Education: First Medical Institute (Leningrad), 1985, MD (internal medicine). Ivanovsky Institute of Virology, 1989, PhD (virology)

Mary Bedell, National Research Service Award Postdoctoral Fellowship, Mammalian Genetics Laboratory. Education: Western Michigan University, 1977, BA (biomedical science). University of Chicago, 1990, PhD (molecular genetics/cell biology)

Ronald Bishop, Chemistry of Carcinogenesis Laboratory. Education: Youngstown State University, 1981, AB (chemistry). West Virginia University School of Medicine, 1990, PhD (biochemistry)

Kristina Borror, Molecular Mechanisms of Carcinogenesis Laboratory. Education: University of Nevada, 1985, BS (biology, chemistry). University of Pittsburgh, 1993, PhD (biochemistry)

Paul Boyer, Molecular Mechanisms of Carcinogenesis Laboratory. Education: Pennsylvania State University, 1980, BS (biochemistry). Michigan State University, 1988, PhD (biochemistry, molecular biology)

Camilynn Brannan, National Neurofibromatosis Foundation Fellowship, Mammalian Genetics Laboratory. Education: Harvey Mudd College, 1985, BS (chemistry). Princeton University, 1990, PhD (molecular biology)

Karen Canella, Chemistry of Carcinogenesis Laboratory. Education: Drexel University, 1980, BS (chemistry). Johns Hopkins University, 1987, PhD (organic chemistry)

Jeffrey Ceci, Leukemia Society of America Postdoctoral Fellowship, Mammalian Genetics Laboratory. Education: Saint Bonaventure University, 1981, BS (biology). University of South Carolina, 1987, PhD (biology)

Mi-Young Chae, Chemistry of Carcinogenesis Laboratory. Education: Seoul National University, 1986, BS (chemistry and education); 1988, MA (science education). Ohio State University, 1993, PhD (chemistry)

Linzhao Cheng, Basic Research Program Postdoctoral Fellowship, Mammalian Genetics Laboratory. Education: University of Science and Technology of China, 1985, BSc (biology). Johns Hopkins University Medical School, 1991, PhD (molecular biology/genetics)

Taesaeng Choi, Molecular Mechanisms of Carcinogenesis Laboratory. Education: Korea University, 1983, BSc (animal science). Tokyo University, 1986, MS (developmental biology); 1992, PhD (developmental biology)

Donna F. Christner, Chemistry of Carcinogenesis Laboratory. Education: Marywood College, 1982, BS (chemistry). University of Maryland at College Park, 1992, PhD (biochemistry)

Vaughn Cleghon, Molecular Mechanisms of Carcinogenesis Laboratory. Education: University of Utah, 1976, BS (psychology); 1985, MBA (business administration). Rutgers, The State University of New Jersey, 1991, PhD (microbiology/molecular genetics)

Janelle Cortner, Molecular Mechanisms of Carcinogenesis Laboratory. Education: University of Washington at Seattle, 1985, BS (biology/chemistry). University of Wisconsin at Madison, 1991, PhD (molecular biology)

Grzegorz Czerwinski, Basic Research Program Postdoctoral Fellowship, Macromolecular Structure Laboratory. Education: Adam Mickiewicz University, 1982, MSc (chemistry); 1992, PhD (chemistry)

Jack Fabian, Molecular Mechanisms of Carcinogenesis Laboratory. Education: Hobart College, 1985, BS (biology). State University of New York (Roswell Park), 1992, PhD (molecular biology)

William Fu, Laboratory of Molecular Virology and Carcinogenesis. Education: Soochow University, 1983, BS (chemistry). West Virginia University, 1991, PhD (biochemistry)

Kenji Fukasawa, Molecular Mechanisms of Carcinogenesis Laboratory. Education: University of Connecticut, 1987, BS (biological sciences). Columbia University, 1989, MS (biological sciences); 1989, MPhil; 1991, PhD (molecular biology)

Shiv Inder Grewal, Laboratory of Eukaryotic Gene Expression. Education: Punjab Agricultural University, 1986, BSc (zoology); 1989, MSc (microbiology). University of Cambridge, 1992, PhD (microbiology)

Leonora Groves, Chemistry of Carcinogenesis Laboratory. Education: West Virginia University, 1989, BA (biology); 1992, PhD (molecular biology)

Finbarr Hayes, Laboratory of Chromosome Biology. Education: University College (Cork), 1984, BSc (microbiology); 1990, PhD (microbiology)

Xi He, Mammalian Genetics Laboratory. Education: Huazhong University of Science and Technology, 1982, BSc (mechanical engineering); 1985, MSc (engineering science). University of California at San Diego, 1992, PhD (biology)

Susan Holbeck, Laboratory of Eukaryotic Gene Expression. Education: University of California at San Diego, 1977, BA (biology). University of Colorado Health Science Center, 1981, MS (microbiology/immunology). University of Washington, 1991, PhD (pathology)

Carolyn Hustad, Mammalian Genetics Laboratory. Education: University of Miami, 1985, BS (chemistry). University of Southern California, 1991, PhD (biochemistry)

Michael Jeffers, Molecular Mechanisms of Carcinogenesis Laboratory. Education: Rutgers University, 1981, BA (microbiology). Columbia University, 1985, MA (science education). New York University, 1993, PhD (biomedical science)

Monica Justice, Mammalian Genetics Laboratory. Education: Fort Hays State University, 1977, BS (medical technology). Kansas State University, 1987, PhD (biology/developmental genetics)

Tadeusz Kochel, Laboratory of Molecular Virology and Carcinogenesis. Education: University of Washington, 1982, BS (chemistry). University of Cincinnati, 1986, MS (biochemistry); 1988, PhD (biochemistry)

Han-Mo Koo, Molecular Mechanisms of Carcinogenesis Laboratory. Education: Yonsei University, 1986, BS (biology). Rutgers — The State University of New Jersey, 1993, PhD (microbiology)

Marina Korneyeva, Molecular Mechanisms of Carcinogenesis Laboratory. Education: USSR Academy of Medical Science, 1979, BS (biochemistry); 1979, MD (biology); 1985, PhD (virology)

Satoshi Kubota, Laboratory of Molecular Virology and Carcinogenesis. Education: Osaka University, 1986, DDS. Kyoto University, 1990, PhD (virology)

Dolores Lana, Molecular Mechanisms of Carcinogenesis Laboratory. Education: University of Maryland, 1978, BS (biology). Virginia-Maryland Regional College of Veterinary Medicine, 1982, MS (immunology); 1987, PhD (immunology)

David Largaespada, Basic Research Program Postdoctoral Fellowship, Mammalian Genetics Laboratory. Education: University of Minnesota, 1986, BS (genetics/cell biology). University of Wisconsin at Madison, 1992, PhD (molecular biology)

Salome La Roche, Laboratory of Eukaryotic Gene Expression. Education: Swiss Federal Institute of Technology, 1987, Diploma (natural sciences); 1991, PhD (natural sciences)

Xiaobei Liao, Mammalian Genetics Laboratory. Education: Beijing University, 1982, BS (biochemistry). University of Connecticut Health Center, 1989, PhD (molecular biology/biochemistry)

A. Jeannine Lincoln, Molecular Mechanisms of Carcinogenesis Laboratory. Education: Olivet College, 1987, BA (biochemistry). Wright State University, 1992, PhD (biomedical sciences)

Jacek Lubkowski, Macromolecular Structure Laboratory. Education: University of Gdansk, 1986, MS (chemistry); 1991, PhD (physical and theoretical chemistry)

Lyudmila Lyakh, Laboratory of Molecular Virology and Carcinogenesis. Education: Minsk State Medical Institute, 1980, MD. Ivanovsky Institute of Virology, 1984, PhD (virology)

Wayne Matten, Molecular Mechanisms of Carcinogenesis Laboratory. Education: University of South Florida, 1976, BS (zoology). University of North Carolina at Chapel Hill, 1990, PhD (biochemistry)

Mark McDougall, Chemistry of Carcinogenesis Laboratory. Education: University of California at Los Angeles, 1981, BS (chemistry). Ohio State University, 1987, PhD (inorganic chemistry)

Barbel Meissner, Molecular Mechanisms of Carcinogenesis Laboratory. Education: University of Heidelberg, 1987, Diploma (biology); 1991, PhD (biology)

Luis Menendez-Arias, Laboratory of Molecular Virology and Carcinogenesis. Education: Complutense University, 1984, BS, MS (biology); 1989, PhD (biology/biochemistry)

Neil R. Michaud, Basic Research Program Postdoctoral Fellowship, Molecular Mechanisms of Carcinogenesis Laboratory. Education: Hobart College, 1986, BS (biology). University of Rochester, 1992, PhD (biology)

Victor Mikhailenko, Chemistry of Carcinogenesis Laboratory. Education: Kiev State University, 1979, BS (biophysics). R.E. Kavetsky Institute for Oncology Problems, 1988, PhD (experimental oncology)

Kristine Miller, Macromolecular Structure Laboratory. Education: Kalamazoo College, 1982, BA (chemistry). University of Notre Dame, 1987, PhD (physical chemistry)

Yury Monczak, Molecular Mechanisms of Carcinogenesis Laboratory. Education: McGill University, 1982, BSc (microbiology/immunology). University of Montreal, 1985, MSc (microbiology/immunology); 1990, PhD (molecular genetics)

Monica Murakami, Molecular Mechanisms of Carcinogenesis Laboratory. Education: University of Maryland, 1986, BS (zoology). Cornell University Graduate School of Medical Sciences, 1992, PhD (molecular biology)

Georgios Nasioulas, Molecular Mechanisms of Carcinogenesis Laboratory. Education: University of Berlin, 1984, BS (biochemistry). German Cancer Research Center, 1989, PhD (biochemistry)

Dwight Nissley, Laboratory of Eukaryotic Gene Expression. Education: Goshen College, 1981, BA (biology). University of Notre Dame, 1990, PhD (microbiology)

Bhavani Pathak, Mammalian Genetics Laboratory. Education: University of Delhi, 1985, BSc (zoology). University of Cincinnati, 1993, PhD (molecular genetics)

William Perry III, Mammalian Genetics Laboratory. Education: University of Tennessee, 1986, BA (college scholars with emphasis in genetics). Columbia University, 1992, PhD (genetics and development)

James Pickel, Mammalian Genetics Laboratory. Education: College of William and Mary, 1975, AB (English literature). University of Alabama at Birmingham, 1991, PhD (microbiology)

Valerian B. Pinto, Molecular Mechanisms of Carcinogenesis Laboratory. Education: St. Xavier's College (Bombay), 1978, BSc (microbiology). Maharaha Sayajirao University (Baroda), 1980, MSc (microbiology). Louisiana State University at Shreveport, 1987, PhD (microbiology/immunology)

Bradford Powell, Laboratory of Chromosome Biology. Education: University of Colorado at Boulder, 1983, BA (biochemistry/molecular biology). University of California at Davis, 1989, PhD (biochemistry)

Chengyi Qian, Chemistry of Carcinogenesis Laboratory. Education: Beijing College of Traditional Chinese Medicine, 1982, BS (pharmacy). Texas Christian University, 1988, MS (chemistry). Baylor University, 1991, PhD (chemistry)

Lyndsay Radnedge, Laboratory of Chromosome Biology. Education: Surrey University, 1982, BSc (microbiology). London University, 1988, PhD (medicine/pharmacy)

Lori Rinckel, Laboratory of Eukaryotic Gene Expression. Education: Xavier University, 1983, BS (biology). University of Illinois, 1988, MS (microbiology); 1992, PhD (microbiology)

Helen Ross, Chemistry of Carcinogenesis Laboratory. Education: University of Nottingham, 1985, BSc (biology); 1990, PhD (biochemistry)

Michael Routledge, Chemistry of Carcinogenesis Laboratory. Education: University of Newcastle upon Tyne, 1986, BSc (cell biology). University of York, 1991, DPhil (biology)

Mary Saltarelli, Molecular Mechanisms of Carcinogenesis Laboratory. Education: Bradley University, 1983, BA (biology). Johns Hopkins University School of Medicine, 1991, PhD (molecular biology)

Brigitte Schmidt, Macromolecular Structure Laboratory. Education: University of Kaiserslautern, 1986, Diploma (chemistry); 1989, PhD (chemistry)

Ralf Schneider, Molecular Mechanisms of Carcinogenesis Laboratory. Education: University of Heidelberg, 1985, Diploma (biology). University of Munich, 1991, PhD (biology)

Gil Sharon, Laboratory of Eukaryotic Gene Expression. Education: Hebrew University of Jerusalem, 1979, BSc (biology); 1981, MSc (genetics); 1989, PhD (biological chemistry)

John Shaughnessy, Mammalian Genetics Laboratory. Education: Christopher Newport College, 1987, BS (biology). University of Maryland, 1993, PhD (molecular biology)

Lisa S.G. Shelton, Molecular Mechanisms of Carcinogenesis Laboratory. Education: University of Maryland at College Park, 1986, BS (microbiology). Uniformed Services University of Health Sciences, 1992, PhD (microbiology)

Jagmohan Singh, Laboratory of Eukaryotic Gene Expression. Education: Indian Institute of Science, 1977, BSc (biochemistry/chemistry); 1978, MSc (biochemistry); 1985, PhD (biochemistry)

Seong K. Song, Molecular Mechanisms of Carcinogenesis Laboratory. Education: Kon-Kuk University, 1982, BA (animal science). University of Texas at Dallas, 1987, MS (molecular biology); 1992, PhD (molecular virology)

Dan Soppet, Foundation for Advanced Cancer Studies Postdoctoral Fellowship, Mammalian Genetics Laboratory. Education: University of California at Santa Cruz, 1978, BA (biology/psychology). Washington University, 1989, PhD (biological sciences)

Eirikur Steingrimssom, Mammalian Genetics Laboratory. Education: University of Iceland, 1985, BS (biology). University of California at Los Angeles, 1992, PhD (biology)

Robert Stephens, Molecular Mechanisms of Carcinogenesis Laboratory. Education: University of Maine, 1978, BS (biochemistry); 1979, MS (biochemistry). Johns Hopkins University, 1991, PhD (molecular biology)

Esta Sterneck, Molecular Mechanisms of Carcinogenesis Laboratory. Education: University of Heidelberg, 1992, PhD (biology)

Jan Szeliga, Chemistry of Carcinogenesis Laboratory. Education: Silesian Technical University, 1976, BS (chemistry). Institute of Oncology (Gliwice), 1990, PhD (chemistry)

Lino Tessarollo, Mammalian Genetics Laboratory. Education: University of Padua, 1987, Diploma (biological sciences); 1989, PhD (biology)

Narmada Thanki, Macromolecular Structure Laboratory. Education: Queen Mary College, University of London, 1985, BSc (chemistry/computer science). Birkbeck College, University of London, 1991, PhD (crystallography)

Genevieve Thon, Laboratory of Eukaryotic Gene Expression. Education: University of Paris, 1984, Maitrise. University of Paris/Pasteur Institute, 1985, DEA

Ilan Tsarfaty, Molecular Mechanisms of Carcinogenesis Laboratory. Education: Tel Aviv University, 1983, BSc (biology); 1985, MSc (microbiology); 1991, PhD (microbiology)

Pantelis Tsoulfas, Mammalian Genetics Laboratory. Education: Pavia Medical School, Italy, 1987, MD

Antonio L. Valentin, Molecular Mechanisms of Carcinogenesis Laboratory. Education: Medicine High School Complutense (Madrid), 1974, BS; 1981, MD. Karolinska Institute, 1992, PhD (virology)

Alfons Valera, Mammalian Genetics Laboratory. Education: Universitat Autonoma de Barcelona, 1989, VMD (veterinary medicine); 1991, MS (biochemistry); 1993, PhD (biochemistry)

Kristine Vogel, Mammalian Genetics Laboratory. Education: Rice University, 1983, BA (biology/anthropology). University of Oregon, 1988, PhD (developmental biology)

Jeannette Whitcomb, Molecular Mechanisms of Carcinogenesis Laboratory. Education: Widener University, 1981, BS (biology). Temple University, 1988, PhD (microbiology/immunology)

Dineli Wickramasinghe, Mammalian Genetics Laboratory. Education: Mount Holyoke College, 1986, AB (biology). Tufts University, 1991, PhD (cell biology)

Simon Williams, Leukemia Society of America Postdoctoral Fellowship, Molecular Mechanisms of Carcinogenesis Laboratory. Education: Trinity College (Dublin), 1983, BA (genetics). State University of New York at Buffalo, 1989, PhD (cellular and molecular biology)

Helen Wilson, Laboratory of Chromosome Biology. Education: Georgia State University, 1982, BS (biology). University of Alabama at Birmingham, 1991, PhD (microbiology)

Sung-Il Yang, Molecular Mechanisms of Carcinogenesis Laboratory. Education: Seoul National University, 1981, BS (pharmacy); 1983, MS (pharmacy). University of Texas Southwestern Medical Center at Dallas, 1992, PhD (pharmacology)

Kazunori Yoshimura, Laboratory of Molecular Virology and Carcinogenesis. Education: Tokyo Metropolitan University, 1979, BSc (chemistry); 1981, MSc (biochemistry/chemistry). Saitama Medical School, 1989, PhD (medicine)

Matthew Young, Laboratory of Molecular Virology and Carcinogenesis. Education: Brooklyn College, 1985, BA (chemistry). University of Maryland, 1990, PhD (biochemistry)

Andrei Zolotukhin, Molecular Mechanisms of Carcinogenesis Laboratory. Education: Moscow State University, 1981, BS (chemistry); 1985, PhD (chemistry)

Summer Student Fellowship Program

Undergraduate and graduate students from colleges and universities in the United States and abroad participate in the Summer Student Fellowship Program for about ten weeks during the summer months. The program is designed to give the students firsthand laboratory experience under the guidance of a senior research scientist. In addition, this year two science teachers participated in the program to gain valuable experience in laboratory techniques that, in turn, could be applied toward the instruction of students in the local schools. The following teacher interns and students were awarded fellowships by the Basic Research Program during the reporting year.

Joanne Austin, from Wake Forest University. Host investigator, Eugene Barsov

Christine Browning, from Hood College. Host investigator, Lyndsay Radnedge

Patricia Carney, from the Rochester Institute of Technology. Host investigator, Barbara Felber

Norman De Rosa, BS, Teacher Intern from Mount Savage High School (on a fellowship provided by the Martin Marietta Corporation through the University of Maryland at Baltimore County) and Graduate of the Governor's Academy for Mathematics, Science and Technology. Host investigator, James Resau

Lawrence Hassner, from the Weizmann Institute of Science, Rehovot, Israel. Host investigator, Christopher Michejda

Susana Hong, from Wesleyan University. Host investigator, Peter Donovan

Ross **Johnston,** from the University of Michigan. Host investigator, Alan Rein

Eric Klineberg, from Rice University. Host investigator, James Resau

Dmitry Kuprash, MS, from the Moscow Institute of Physics and Technology, Moscow, Russia. Host investigator, Nancy Rice

Ilsa Mae Lambert, from the University of California at Santa Cruz. Host investigator, Lori Rinckel

Jeannette Liu, from Brown University. Host investigator, Ann Abeles

Bryan Main, from Frostburg State University. Host investigator, Alexander Wlodawer

Felice Mirsky, from Vassar College. Host investigator, Michael Routledge

Ashmi Patel, from the University of California at Berkeley. Host investigator, Esta Sterneck

Julie Pavlovsky, from Frederick Community College. Host investigator, Ronald Rubin

Preetha Pillai, from the University of Virginia. Host investigator, Stephen Oroszlan

Shelley Smith, MS, Teacher Intern from Frederick High School. Host investigator, Sharon Moore

Lisa Taneyhill, from Western Maryland College. Host investigator, Marilyn Kroeger-Koepke

Tung Tran, BS, from the University of Texas Southwestern Medical School. Host investigator, Mark Baer

Publications

Adachi Y, Copeland TD, Hatanaka M, Oroszlan, S: Nucleolar targeting signal of Rex protein of human T-cell leukemia virus type I specifically binds to nucleolar shuttle protein B-23. J Biol Chem 268:13930, 1993

Adachi Y, Copeland TD, Takahashi C, Nosaka T, Ahmed A, Oroszlan S, Hatanaka M: Phosphorylation of the Rex protein of human T-cell leukemia virus type-1. J Biol Chem 267:21977, 1992

Allen RC, Armitage RJ, Conley ME, Rosenblatt H, Jenkins NA, Copeland NG, Bedell MA, Edelhoff S, Disteche CM, Simoneaux DK, Fanslow WC, Belmont J, Spriggs MK: CD40 ligand gene defects responsible for X-linked hyper-IgM syndrome. Science 259:990, 1993

Arcangioli B, Klar AJS: The programme of cell type determination in fission yeast. In: The Molecular Genetics of Differentiation, Springer-Verlag, p 75, 1992

Barker CS, Bear SE, Keler T, Copeland NG, Gilbert DG, Jenkins NA, Yeung RS, Tsichlis PN: Activation of the prolactin receptor gene by promoter insertion in a Moloney leukemia virus (MoMuLV) induced rat thymoma. J Virol 66:6763, 1992

Behan DP, Potter E, Lewis KA, Jenkins NA, Copeland NG, Lowry PJ, Vale WW: Cloning and structure of the human corticotrophin releasing factor-binding protein gene. Genomics 16:63, 1993

Beisel CE, Edwards JF, Dunn LL, Rice NR: Analysis of multiple mRNAs from pathogenic equine infectious anemia virus (EIAV) in an acutely infected horse reveals a novel protein, Ttm, derived from the carboxy terminus of the EIAV transmembrane protein. J Virol 67:832, 1993

Bishop RE, Dunn LL, Pauly GT, Dolan ME, Moschel RC: The role of O^6-alkylguanine-DNA alkyltransferase in protecting Rat4 cells against the mutagenic effects of O^6-substituted guanine residues incorporated in codon 12 of the H-*ras* gene. Carcinogenesis 14:593, 1993

Blaha I, Tozser J, Kim Y, Copeland TD, Oroszlan S: Solid phase synthesis of the proteinase of bovine leukemia virus. FEBS Lett 309:389, 1992

Boyer PL, Colmenares C, Stavnezer E, Hughes SH: Sequence and biological activity of chicken *sno*N cDNA clones. Oncogene 8:457, 1993

Boyer PL, Currens MJ, McMahon JB, Boyd MR, Hughes SH: Analysis of nonnucleoside drug-resistant variants of human immunodeficiency virus type 1 reverse transcriptase. J Virol 67:2412, 1993

Boyer PL, Ferris AL, Hughes SH: Mutational analysis of the fingers domain of human immunodeficiency virus type 1 reverse transcriptase. J Virol 66:7533, 1992

Brannan CI, Bedell MA, Resnick JL, Eppig JJ, Handel MA, Williams DE, Lyman SD, Donovan PJ, Jenkins NA, Copeland NG: Developmental abnormalities in *Steel*^{17H} mice result from a splicing defect in the steel factor cytoplasmic tail. Genes Dev 6:1832, 1992

Brannan CI, Gilbert DJ, Ceci JD, Matsuda Y, Chapman VM, Mercer JA, Eisen H, Johnston LA, Copeland NG, Jenkins NA: An interspecific linkage map of mouse chromosome 15 positioned with respect to the centromere. Genomics 13:1075, 1992

Brown S: Engineered iron oxide-adhesion mutants of the *Escherichia coli* phage λ receptor. Proc Natl Acad Sci USA 89:8651, 1992

Cameron CE, Grinde B, Jacques P, Jentoft J, Leis J, Wlodawer A, Weber IT: Comparison of the substrate-binding pockets of the Rous sarcoma virus and human immunodeficiency virus type I proteases. J Biol Chem 268:11711, 1993

Cameron CE, Grinde B, Jentoft J, Leis J, Weber IT, Copeland TD, Wlodawer A: Mechanism of inhibition of the retroviral protease by a Rous sarcoma virus peptide substrate representing the cleavage site between the *gag* p2 and p10 proteins. J Biol Chem 267:23735, 1992

Canella KA, Peltonen K, Yagi H, Jerina DM, Dipple A: Identification of individual benzo[*c*]phenanthrene dihydrodiol epoxide–DNA adducts by the ^{32}P-postlabeling assay. Chem Res Toxicol 5:685, 1992

Cattanach BM, Barr JA, Evans EP, Burtenshaw M, Beechey CV, Leff SE, Brannan CI, Copeland NG, Jenkins NA, Jones J: A candidate mouse model for Prader Willi syndrome which shows an absence of Snrpn expression. Nature Genet 2:270, 1992

Ceci JD, Lusis AJ: Mouse chromosome 8. Mammal Genome 3:S121, 1992

Chen J-M, Zhang Y-P, Moschel RC, Ikenaga M: Depletion of O^6-methylguanine–DNA methyltransferase and potentiation of 1,3-bis(2-chloroethyl)-1-nitrosourea antitumor activity by O^6-benzylguanine *in vitro*. Carcinogenesis 14:1057, 1993

Chen J-M, Zhang Y-P, Sui J-L, Moschel RC, Ikenaga M: Modulation of O^6-methylguanine-DNA methyltransferase-mediated 1-(4-amino-2-methyl-5-pyrimidinyl) methyl-3-(2-chloroethyl)-3-nitrosourea resistance by O^6-benzylguanine *in vitro* and *in vivo*. Anticancer Res 13:801, 1993

Cole TJ, Copeland NG, Gilbert DJ, Jenkins NA, Schutz G, Ruppert S: The mouse *CREB* (cAMP responsive element binding protein) gene: Structure, promoter analysis, and chromosomal localization. Genomics 13:974, 1992

Constantoulakis P, Campbell M, Felber BK, Nasioulas G, Afonina E, Pavlakis GN: Inhibition of Rev-mediated HIV-1 expression by an RNA binding protein encoded by the interferon-inducible 9-27 gene. Science 259:1314, 1993

Copeland NG: Project reading. Review of *Mammalian Genome: Official Journal of the International Mammalian Genome Society*. Nature 359:440, 1992

Copeland NG, Gilbert DJ, Chretien M, Seidah NG, Jenkins NA: Regional localization of three convertases, PC1 (*Nec-1*), PC2 (*Nec-2*), and furin (*Fur*), on mouse chromosomes. Genomics 13:1356, 1992

Copeland NG, Gilbert DJ, Li K, Sawamura D, Giudice GJ, Chu M-L, Jenkins NA, Uitto J: Chromosomal localization of mouse bullous pemphigoid antigens, BPAG1 and BPAG2: Identification of a new region of homology between mouse and human chromosomes. Genomics 15:180, 1993

Court D: RNA processing and degradation by RNase III. In: Control of Messenger RNA Stability (Brawerman G, Belasco J, eds) Academic Press, p 71, 1993

Daar I, Yew N, Vande Woude GF: Inhibition of *mos*-induced oocyte maturation by protein kinase A. J Cell Biol 120:1197, 1993

Danishefsky AT, Burton LE, Rubin JR: Crystallization and preliminary characterization of three crystal forms of human recombinant transforming growth factor-α. Biochem Biophys Res Comm 187:146, 1992

Danishefsky AT, Housset D, Kim K-S, Tao F, Fuchs J, Woodward C, Wlodawer A: Crevice-forming mutants in the rigid core of bovine pancreatic trypsin inhibitor: Crystal structures of F22A, Y23A, N43G, and F45A. Protein Sci 2:577, 1993

Derr LK, Strathern JN: A role for reverse transcripts in gene conversion. Nature 361:170, 1993

Dickson B, Sprenger F, Morrison D, Hafen E: Raf functions downstream of Ras1 in the Sevenless signal transduction pathway. Nature 360:600, 1992

Disteche CM, Brannan CI, Larsen A, Adler DA, Schorderet DF, Gearing D, Copeland NG, Jenkins NA, Park LS: Evolution of the pseudoautosomal region: the GM-CSF receptor α subunit gene is on mouse chromosome 19. Nature Genet 1:333, 1992

Dolan ME, Pegg AE, Biser ND, Moschel RC, English HF: Effect of O^6-benzylguanine on the response to 1,3-bis(2-chloroethyl)-1-nitrosourea in the dunning R3327G model of prostatic cancer. Cancer Chemother Pharmacol 32:221, 1993

Dolan ME, Pegg AE, Moschel RC, Grindey GB: Effect of O^6-benzylguanine on the sensitivity of human colon tumor xenografts to 1,3-BIS(2-chloroethyl)-1-nitrosourea (BCNU). Biochem Pharmacol 46:285, 1993

Faletto DL, Kaplan DR, Halverson DO, Rosen EM, Vande Woude GF: Signal transduction in c-*met* mediated motogenesis. In: Hepatocyte Growth Factor-Scatter Factor (HGF-SF) and the C-*Met* Receptor (Goldberg ID, Rosen EM, eds) Birkhauser Verlag, p 107, 1993

Feng YX, Yuan H, Rein A, Levin JG: A bipartite signal for readthrough suppression in murine leukemia virus mRNA: an eight-nucleotide purine-rich sequence immediately downstream of the *gag* termination codon followed by an RNA pseudoknot. J Virol 66:5127, 1992

Foroni L, Boehm T, White L, Forster A, Sherrington P, Liao XB, Brannan CI, Jenkins NA, Copeland NG, Rabbitts TH: The rhombotin gene family encode related LIM-domain proteins whose differing expression suggests multiple roles in mouse development. J Mol Biol 226:747, 1992

Forry-Schaudies S, Hughes SH: Structure, evolution, and expression of the tropomyosin gene family. In: Genome Research in Molecular Medicine and Virology, Academic Press, p 83, 1993

Franco del Amo F, Gendron-Maguire M, Swiatek PJ, Jenkins NA, Copeland NG, Gridley T: Cloning, analysis and chromosomal localization of *Motch*, a mouse homolog of *Drosophila Notch*. Genomics 15:259, 1993

Frew T, Whitaker B, Salter D, Payne W, Hughes SH, Yamamoto H, Takeuchi T, Brumbaugh J: The expression of mouse tyrosinase in chick cells in vitro and in vivo when controlled by a constitutive promoter. Pigment Cell Res 5:312, 1992

Friedman HS, Dolan ME, Moschel RC, Pegg AE, Felker GM, Rich J, Bigner DD, Schold SC Jr: Enhancement of nitrosourea activity in medulloblastoma and glioblastoma multiforme. J Natl Cancer Inst 84:1926, 1992

Friezner Degen SJ, Gilbert DJ, Jenkins NA, Copeland NG: Assignment of the gene coding for hepatocyte growth factor-like protein to mouse chromosome 9. Genomics 13:1368, 1992

Fu W, Rein A: Maturation of dimeric viral RNA of Moloney murine leukemia virus. J Virol 67:5443, 1993

Garfinkel DJ: Retroelements in microorganisms. In: The Retroviridae, Vol. 1 (Levy JA, ed) Plenum Press, p 107, 1992

Gartner J, Kearns W, Rosenberg C, Pearson P, Copeland NG, Gilbert DJ, Jenkins NA, Valle D: Localization of the 70-kDa peroxisomal membrane protein to human 1p21-p22 and mouse 3. Genomics 15:412, 1993

Ghosh P, Tan TH, Rice NR, Sica A, Young HA: The IL-2 CD28-responsive complex contains at least three members of the NF-κB family: c-Rel, p50, and p65. Proc Natl Acad Sci USA 90:1696, 1993

Gilbert DJ, Neumann PE, Taylor BA, Jenkins NA, Copeland NG: Susceptibility of AKXD recombinant inbred mouse strains to lymphomas. J Virol 67:2083, 1993

Goldsborough AS, Healy LE, Copeland NG, Gilbert DJ, Jenkins NA, Willison KR, Ashworth A: Cloning, chromosomal localization and expression pattern of the POU domain gene *Oct-11*. Nucleic Acids Res 21:127, 1993

Gorman DM, Itoh N, Jenkins NA, Gilbert DJ, Copeland NG, Miyajima A: Chromosomal localization and organization of the murine genes encoding the β subunits (AIC2A and AIC2B) of the interleukin 3, granulocyte/macrophage colony-stimulating factor, and interleukin 5 receptors. J Biol Chem 267:15842, 1992

Gorski DH, LePage DF, Patel CV, Copeland NG, Jenkins NA, Walsh K: Molecular cloning of a diverged homeobox gene that is rapidly down-regulated during the G0/G1 transition in vascular smooth muscle cells. Genes Dev 13:3722, 1993

Gribskov M: The language metaphor in sequence analysis. Computers Chem 16:85, 1992

Gulnik S, Baldwin ET, Tarasova N, Erickson J: Human liver cathepsin D. J Mol Biol 227:265, 1992

Gustchina A, Weber IT, Wlodawer A: Molecular modeling of the HIV-2 protease. In: Structure and Function of the Aspartic Proteinases: Genetics, Structure, and Mechanisms (Dunn B, ed) Plenum, p 549, 1992

Hayes F, Davis MA, Austin SJ: Fine-structure analysis of the P7 plasmid partition site. J Bacteriol 175:3443, 1993

Hempstead BL, Rabin SL, Kaplan L, Reid S, Parada LF, Kaplan DR: Overexpression of the trk tyrosine kinase rapidly accelerates nerve growth factor-induced differentiation. Neuron 9:1, 1992

Herskowitz I, Rine J, Strathern J: Mating-type determination and mating-type interconversion in *Saccharomyces cerevisiae*. In: The Molecular and Cellular Biology of the Yeast *Saccharomyces*, Vol II, Gene Expression (Broach JR, Jones EW, Pringle JR, eds) Cold Spring Harbor Laboratory Press, p 583, 1992

Heusel JW, Scarpati EM, Jenkins NA, Gilbert DJ, Copeland NG, Shapiro SD, Ley TJ: Molecular cloning, chromosomal location, and tissue-specific expression of the murine cathepsin G gene. Blood 81:1614, 1993

Hizi A, Tal R, Shaharabany M, Currens MJ, Boyd MR, Hughes SH, McMahon JB: Specific inhibition of the reverse transcriptase of human immunodeficiency virus type-1 and the chimeric enzymes of human immunodeficiency viruses type-1 and type-2 by the novel nonnucleoside inhibitor calanolide A. Antimicrob Agents Chemother 37:1037, 1993

Holtzman DM, Li Y, Parada LF, Kinsman S, Chen C-K, Valletta JS, Zhou J, Long JB, Mobley WC: $p140^{trk}$ mRNA marks NGF-responsive forebrain neurons: Evidence that *trk* gene expression is induced by NGF. Neuron 9:465, 1992

Hruszkewycz AM, Canella KA, Peltonen K, Kotrappa L, Dipple A: DNA polymerase action on benzo[α]pyrene–DNA adducts. Carcinogenesis 13:2347, 1992

Jacobo-Molina A, Ding J, Nanni RG, Clark AD, Lu X, Tantillo C, Williams RL, Kamer G, Ferris AL, Clark P, Hizi A, Hughes SH, Arnold E: Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 Å resolution shows bent DNA. Proc Natl Acad Sci USA 90: 6320, 1993

Jiang W, Sadler PM, Jenkins NA, Gilbert DJ, Copeland NG, Bond JS: Tissue-specific expression and chromosomal localization of the α subunit of mouse meprin A. J Biol Chem 268:10380, 1993

Johnson PF, Sterneck E, Williams SC: Activation domains of transcriptional regulatory proteins. J Nutr Biochem 4:386, 1993

Justice MJ, Gilbert DJ, Kinzler KW, Vogelstein B, Buchberg AM, Ceci JD, Matsuda Y, Chapman VM, Patriotis C, Makris A, Tsichlis PN, Jenkins NA, Copeland NG: A molecular genetic linkage map of mouse chromosome 18 reveals extensive linkage conservation with human chromosomes 5 and 18. Genomics 13:1281, 1992

Justice MJ, Stephenson DA: Mouse chromosome 13. Mammal Genome 3:S195, 1992

Kahn T, Friesl H, Copeland NG, Gilbert DJ, Jenkins NA, Gissmann L, Kramer J, zur Hausen H: Molecular cloning, analysis and chromosomal localization of a mouse genomic sequence related to the human papillomavirus type 18 E5 region. Mol Carcinogenesis 6:88, 1992

Kaplan DR, Perkins A, Morrison DK: Signal transduction by receptor tyrosine kinases. In: Oncogenes and Tumor Suppressor Genes in Human Malignancies (Benz CC, Liu ET, eds) Kluwer Academic Publishers, p 265, 1993

Karls U, Muller U, Gilbert DJ, Copeland NG, Jenkins NA, Harbers K: Structure, expression, and chromosome location of the gene for the β subunit of brain-specific Ca^{2+}/calmodulin-dependent protein kinase II identified by transgene integration in an embryonic lethal mouse mutant. Mol Cell Biol 12:3644, 1992

Kashman Y, Gustafson KR, Fuller RW, Cardellina JH, McMahon JB, Currens MJ, Buckheit RW Jr, Hughes SH, Cragg GM, Boyd MR: The calanolides, a novel HIV-inhibitory class of coumarin derivatives from the tropical rainforest tree, *Calophyllum lanigerum.* J Med Chem 35, 2735, 1992

Kasprzak K, North SL, Hernandez L: Reversal by nickel(II) of inhibitory effects of some scavengers of active oxygen species upon hydroxylation of 2′-deoxyguanosine in vitro. Chem Biol Interactions 84:11, 1992

Kawakami K, Shafer BK, Garfinkel DJ, Strathern JN, Nakamura Y: Ty element-induced temperature-sensitive mutations of *Saccharomyces cerevisiae.* Genetics 131:821, 1992

Kim K-S, Tao F, Fuchs J, Danishefsky AT, Housset D, Wlodawer A, Woodward C: Crevice-forming mutants of bovine pancreatic trypsin inhibitor: Stability changes and new hydrophobic surface. Protein Sci 2:588, 1993

Kingsley DM, Bland AE, Grubber JM, Marker PC, Russell LB, Copeland NG, Jenkins NA: The mouse short ear skeletal morphogenesis locus is associated with defects in a bone morphogenetic member of the TGF-β superfamily. Cell 71:399, 1992

Klar AJS: Molecular genetics of fission yeast cell type: Mating type and mating-type interconversion. In: The Molecular and Cellular Biology of the Yeast *Saccharomyces*, Vol II, Gene Expression (Broach J, Jones E, Pringle J, eds) Cold Spring Harbor Laboratory Press, p 745, 1992

Kmiecik TE, Keller JR, Rosen E, Vande Woude GF: Hepatocyte growth factor is a synergistic factor for the growth of hematopoietic progenitor cells. Blood 80:2454, 1992

Knusel B, Kaplan DR, Winslow JW, Rosenthal A, Burton LE, Beck KD, Rabin S, Nikolics K, Hefti F: K-252b selectively potentiates cellular actions and *trk* tyrosine phosphorylation mediated by neurotrophin-3. J Neurochem 59:715, 1992

Knusel B, Rabin S, Widmer HR, Hefti F, Kaplan DR: Neurotrophin-induced *trk* receptor phosphorylation and cholinergic neuron response in primary cultures of embryonic rat brain neurons. Dev Neurosci 3:885, 1992

Kolbanovskaya EY, Sathyanarayana BK, Wlodawer A, Karpeisky MY: Intramolecular interactions in pancreatic ribonucleases. Protein Sci 1:1050, 1992

Komuro I, Schalling M, Jahn L, Bodmer R, Jenkins NA, Copeland NG, Izumo S: *Gtx*: A novel murine homeobox-containing gene, expressed specifically in glial cells of the brain and germ cells of testis, has a transcriptional repressor activity *in vitro* for a serum-inducible promoter. EMBO J 12:1387, 1993

Kotler M, Arad G, Hughes SH: Human immunodeficiency virus type 1 *gag*-protease fusion proteins are enzymatically active. J Virol 66:6781, 1992

Leff SE, Brannan CI, Reed ML, Ozcelik T, Francke U, Copeland NG, Jenkins NA: Maternal imprinting of the mouse *Snrpn* gene and conserved linkage homology with Prader-Willi syndrome region of humans. Nature Genet 2:259, 1992

Levin JG, Hatfield DL, Oroszlan S, Rein A: Mechanisms of translational suppression used in the biosynthesis of reverse transcriptase. In: Reverse Transcriptase (Goff S, Skalka AM, eds) Cold Spring Harbor Laboratory Press, p 5, 1993

Li K, Christiano AM, Copeland NG, Gilbert DJ, Chu M-L, Jenkins NA, Uitto J: cDNA cloning and chromosomal mapping of the mouse type VII collagen gene (*Col7a1*): Evidence for rapid evolutionary divergence of the gene. Genomics 16:733, 1993

Lin X-L, Lin Y-Z, Koelsch G, Gutschina A, Wlodawer A, Tang J: Enzymic activities of two-chain pepsinogen, two-chain pepsin, and the amino-terminal lobe of pepsinogen. J Biol Chem 267:17257, 1992

Mano H, Mano K, Tang B, Kohler M, Yi T, Gilbert DJ, Jenkins NA, Copeland NG, Ihle JI: Expression of a novel form of Tec kinase in hematopoietic cells and mapping of the gene to chromosome 5 near Kit. Oncogene 8:417, 1993

McGill CB, Shafer BK, Derr LK, Strathern JN: Recombination initiated by double-strand breaks. Curr Genet 23:305, 1993

McKenzie AN, Li X, Largaespada DA, Sato A, Kaneda A, Zurawski S, Doyle E, Milatovich A, Francke U, Copeland NG, Jenkins NA, Zurawski G: Structural comparison and chromosomal localization of the human and mouse interleukin-13 genes. J Immunol 150:5436, 1993

Menendez-Arias L, Young M, Oroszlan S: Purification and characterization of the mouse mammary tumor virus protease expressed in *Escherichia coli.* J Biol Chem 267:24134, 1992

Michejda M, Peters SM, Bacher J, Hernandez LF, Bellanti JA: Intrauterine xenotransplantation of bone marrow: stem cells in nonhuman primates. Transplantation 54:1, 1992

Miller M, Mohana Rao JK, Wlodawer A, Gribskov M: A left-handed crossover involved in amidohydrolase catalysis: crystal structure of *Erwinia chrysanthemi* L-asparaginase with bound L-aspartate. FEBS Lett 328:275, 1992

Miyatani S, Copeland NG, Gilbert DJ, Jenkins NA, Takeichi M: Genomic structure and chromosomal mapping of the mouse N-cadherin gene. Proc Natl Acad Sci USA 89:8443, 1992

Moschel RC, McDougall MG, Dolan ME, Stine L, Pegg AE: Structural features of substituted purine derivatives compatible with depletion of human O^6-alkylguanine-DNA alkyltransferase. J Med Chem 35:4486, 1992

Nam SH, Copeland TD, Hatanaka M, Oroszlan S: Characterization of ribosomal frameshifting for expression of *pol* gene products of human T-cell leukemia virus type I. J Virol 67:196, 1993

Nett MA, Cerretti CP, Berson DR, Seavitt J, Gilbert DJ, Jenkins NA, Copeland NG, Black RA, Chaplin DD: Molecular cloning of the murine Interleukin-1β converting enzyme cDNA. J Immunol 149:3254, 1992

O'Brien SJ, Womack JE, Lyons LA, Moore KJ, Jenkins NA, Copeland NG: Anchored reference loci for comparative genome mapping in mammals. Science 3:103, 1993

Okamoto K, Wakamiya M, Noji S, Koyama E, Taniguchi S, Takemura R, Copeland NG, Gilbert DJ, Jenkins NA, Muramatsu M, Hamada H: A novel class of murine POU gene predominantly expressed in central nervous system. J Biol Chem 268:7449, 1993

Okazaki T, Yoshida BN, Avraham KB, Wang H, Wuenschell CW, Jenkins NA, Copeland NG, Anderson DJ, Mori N: Molecular diversity of the SCG10/stathmin gene family in the mouse. J Biol Chem 268:7449, 1993

Oroszlan S, Tozser J, Weber IT: The proteinases of bovine leukaemia and equine infectious anaemia viruses. Int Antiviral News 1:22, 1993

Ott DE, Keller J, Sill K, Rein A: Phenotypes of murine leukemia virus-induced tumors: Influence of 3′ viral coding sequences. J Virol 66:6107, 1992

Parada LF, Tsoulfas P, Tessarollo L, Blair J, Reid SW, Soppet D: The *Trk* family of tyrosine kinases: Receptors for NGF-related neurotrophins. Cold Spring Harb Symp Quant Biol 57:43, 1992

Paules RS, Resnick J, Kasenally AB, Ernst MK, Donovan P, Vande Woude GF: Characterization of activated and normal mouse *Mos* gene in murine 3T3 cells. Oncogene 7:2489, 1992

Peltonen K, Canella K, Dipple A: High-performance liquid chromatographic separation of purine deoxyribonucleoside monophosphate–benzo[*a*]pyrene adducts. J Chromatogr 623:247, 1992

Perkins AS, Vande Woude GF: Principles of molecular cell biology of cancer: Oncogenes. In: Cancer: Principles and Practice of Oncology, 4th Edition (DeVita VT Jr, Hellman S, Rosenberg SA, eds) Lippincott, p 3, 1993

Porter BE, Justice MJ, Copeland NG, Jenkins NA, Hunter DD, Merlie JP, Sanes JR: S-laminin: Mapping to mouse chromosome 9 and expression in the linked mutants tippy and ducky. Genomics 16:278, 1993

Posada J, Yew N, Ahn NG, Vande Woude GF, Cooper JA: Mos stimulates MAP kinase in *Xenopus* oocytes and activates a MAP kinase kinase in vitro. Mol Cell Biol 13:2546, 1993

Rabin SJ, Cleghon V, Kaplan DR: SNT, a differentiation-specific target of neurotrophic factor-induced tyrosine kinase activity in neurons and PC12 cells. Mol Cell Biol 13:2203, 1993

Resnick JL, Bixler LS, Cheng L, Donovan PJ: Long-term proliferation of mouse primordial germ cells in culture. Nature 359:550, 1992

Rice NR, MacKichan ML, Israel A: The precursor of NF-κB P50 has IκB-like functions. Cell 71:243, 1992

Rong S, Bodescot M, Blair D, Dunn J, Nakamura T, Mizuno K, Park M, Chan A, Aaronson S, Vande Woude GF: Tumorigenicity of the *met* proto-oncogene and the gene for hepatocyte growth factor. Mol Cell Biol 12:5152, 1992

Rong S, Oskarsson M, Faletto D, Tsarfaty I, Resau J, Nakamura T, Rosen E, Hopkins R, Vande Woude GF: Tumorigenesis induced by co-expression of human hepatocyte growth factor and the human *Met* protooncogene leads to high levels of expression of the ligand and receptor. Cell Growth Diff 4:563, 1993

Ross H, Bigger CAH, Yagi H, Jerina DM, Dipple A: Sequence specificity in the interaction of the four stereoisomeric benzo[*c*]phenanthrene dihydrodiol epoxides with the *supF* gene. Cancer Res 53:1273, 1993

Routledge MN, Wink DA, Keefer LK, Dipple A: Mutations induced by saturated aqueous nitric oxide in the pSP189 *supF* gene in human Ad293 and *E. coli* MBM7070 cells. Carcinogenesis 14:1251, 1993

Rouzer CA, Thompson EJ, Skinner TL, Heavner PA, Bartolini WP, Mitchell K, Kurz E, Smith RH Jr, Michejda CJ: An unexpected pathway for the metabolic degradation of 1,3-dialkyl-3-acyltriazenes. Biochem Pharmacol 46:165, 1993

Safadi M, Bindra DS, Williams T, Moschel RC, Stella VJ: Kinetics and mechanism of the acid-catalyzed hydrolysis of O^6-benzylguanine. Int J Pharmaceut 90:239, 1993

Samudzi CT, Fivash MJ: Cluster analysis of the biological macromolecule crystallization database. J Crystal Growth 123:47, 1992

Samudzi CT, Nguyen NY, Rubin JR: Crystallization and preliminary X-ray diffraction studies of dogfish C-reactive protein. Proteins: Structure, Function, and Genetics 15:100, 1993

Sawyer TK, Staples DJ, Liu L, Tomasselli AG, Hui JO, O'Connell K, Schostarez H, Hester JB, Moon J, Howe WJ, Smith CW, Decamp DL, Craik CS, Dunn BM, Lowther WT, Harris J, Poorman RA, Wlodawer A, Jaskolski M, Heinrikson RL: HIV protease (HIV PR) inhibitor structure-activity-selectivity, and active site molecular modeling of high affinity Leu [$CH(OH)CH_2$]Val modified viral and nonviral substrate analogs. Int J Peptide Protein Res 40:274, 1992

Schulz N, Propst F, Rosenberg MM, Linnoila RI, Paules RS, Schulte D, Vande Woude GF: Patterns of neoplasia in c-*mos* transgenic mice and their relevance to multiple endocrine neoplasia. Henry Ford Hosp Med J 40:307, 1992

Schwartz S, Campbell M, Felber BK, Pavlakis GN: Mutational inactivation of an inhibitory sequence in HIV-1 results in Rev-independent *gag* expression. J Virol 66:7176, 1992

Singh J, Klar AJS: DNA polymerase-α is essential for mating-type switching in fission yeast. Nature 361:271, 1993

Sjolin L, Wlodawer A, Bernqvist G, Holm P, Loth K, Malmstron H, Kaar J, Svensson LA, Dill J, Gilliland GL: Protein crystals grown on board MASER 3 extend the ribonuclease A structure to 1.06 Å resolution. In: European Space Agency SP-1132, Vol 2, p 92, 1992

Smith RH Jr, Pruski B, Day CS, Pfaltzgraff TD, Michejda CJ: Formation of aliphatic azimines in an unexpected reaction. Tetrahedron Lett 33:4683, 1992

Smith RH Jr, Wladkowski BD, Herling JA, Pfaltzgraff TD, Taylor JE, Thompson EJ, Pruski B, Klose JR, Michejda CJ: Novel triazenes and triazolines from the base-catalyzed hydrolysis of 1,3-dialkyl-3-acyltriazenes. J Org Chem 57:6448, 1992

Smith RH Jr, Wladkowski BD, Taylor JE, Thompson EJ, Pruski B, Klose JR, Andrews AW, Michejda CJ: Acid-catalyzed decomposition of 1-alkyltriazolines: A mechanistic study. J Org Chem 58:2097, 1993

Soltoff SP, Rabin SL, Cantley LC, Kaplan DR: Nerve growth factor promotes the activation of phosphatidylinositol 3-kinase and its association with the *trk* tyrosine kinase. J Biol Chem 267:17472, 1992

Sprenger F, Trosclair MM, Morrison DK: Biochemical analysis of Torso and D-Raf during *Drosophila* embryogenesis: Implications for terminal signal transduction. Mol Cell Biol 13:1163, 1993

Stanton BR, Perkins AS, Tessarollo L, Sassoon DA, Parada LF: Loss of N-*myc* function results in embryonic lethality and failure of the epithelial component of the embryo to develop. Genes Dev 6:2235, 1992

Stephens RM, Sithanandam G, Copeland TD, Kaplan DR, Rapp UR, Morrison DK: 95-Kilodalton B-Raf serine/threonine kinase: Identification of the protein and its major autophosphorylation site. Mol Cell Biol 12:3733, 1992

Strathern JN: Recent writings on yeast recombination. Curr Opin Genet Dev 2:691, 1992

Sutrave P, Leferovich J, Kelly AM, Hughes SH: *c-ski* expression can increase the skeletal musculature of transgenic mice. In: Neuromuscular Development and Disease (Kelly AM, Blau HM, eds) Raven Press, p 107, 1992

Swain A, Gustchina A, Wlodawer A: Comparison of three inhibitor complexes of human immunodeficiency virus protease. In: Structure and Function of the Aspartic Proteinases: Genetics, Structure, and Mechanisms (Dunn B, ed) Plenum, p 433, 1992

Swain AL, Jaskolski M, Housset D, Rao JKM, Wlodawer A: Crystal structure of *Escherichia coli* L-asparaginase, an enzyme used in cancer therapy. Proc Natl Acad Sci USA 90:1474, 1993

Tan GT, Wichramasinghe A, Verma S, Singh R, Hughes SH, Pezzuto JM, Baba M, Mohan P: Potential anti-AIDS napthalenesulfonic acid derivatives. Synthesis and inhibition of HIV-1 induced cytopathogenesis and HIV-1 and HIV-2 reverse transcriptase activities. J Med Chem 35:4846, 1992

Tan TH, Huang GP, Sica A, Ghosh P, Young HA, Longo DL, Rice NR: κB site-dependent activation of the interleukin 2 receptor α chain gene promoter by human c-*rel*. Mol Cell Biol 12:4067, 1992

Thanki N, Rao JKM, Foundling SI, Howe WJ, Moon JB, Hui JO, Tomasselli AG, Heinrikson RL, Thaisrivongs S, Wlodawer AL: Crystal structure of a complex of HIV-1 protease with a dihydroxyethylene-containing inhibitor: Comparisons with molecular modeling. Protein Sci 1:1061, 1992

Tozser J, Friedman D, Weber IT, Blaha I, Oroszlan S: Studies on the substrate specificity of the proteinase of equine infectious anemia virus using oligopeptide substrates. Biochemistry 32:3347, 1993

Tsarfaty I, Resau JH, Rulong S, Keydar I, Faletto DL, Vande Woude GF: The *met* proto-oncogene receptor and lumen formation. Science 257:1258, 1992

Tsoulfas P, Soppet D, Escandon E, Tessarollo L, Mendoza-Ramirez J-L, Rosenthal A, Nikolics K, Parada LF: The rat *trk*C locus encodes multiple neurogenic receptors that exhibit differential response to neurotrophin-3 in PC12 cells. Neuron 10:975, 1993

Tsui FWL, Tsui H-W, Mok S, Mlinaric I, Copeland NG, Gilbert DJ, Jenkins NA, Siminovitch KA: Molecular characterization and mapping of murine genes encoding three members of the stefin family of cysteine proteinase inhibitors. Genomics 15:507, 1993

Vande Woude S, Vande Woude GF: Principles of molecular and cellular biology of cancer: Introduction to methods in molecular biology. In: Cancer: Principles and Practice of Oncology, 4th Edition (DeVita VT Jr, Hellman S, Rosenberg SA, eds) Lippincott, p 35, 1993

Weber IT, Tozser J, Wu J, Friedman D, Oroszlan S: Molecular model of equine infectious anemia virus proteinase and kinetic measurements for peptide substrates with single amino acid substitutions. Biochemistry 32:3354, 1993

Weinstock KG, Strathern JN: Molecular genetics in *Saccharomyces kluyveri*: The *HIS3* homolog and its use as a selectable marker gene in *S. kluyveri* and *Saccharomyces cerevisiae*. Yeast 9:351, 1993

Whitcomb JM, Hughes SH: Retroviral reverse transcription and integration: Progress and problems. Annu Rev Cell Biol 8:275, 1992

Wilson JB, Ferguson MWJ, Jenkins NA, Lock LF, Copeland NG, Levine AJ: Transgenic mouse model of X-linked cleft palate. Cell Growth Diff 4:67, 1992

Wlodawer A: Inhibitor complexes of the HIV protease: a target for drug design. In: Proceedings of the Robert A. Welch Foundation Conference on Chemical Research XXXV, p 71, 1992

Wlodawer A: Another piece of the HIV puzzle falls into place. Science 256:1766, 1992

Wlodawer A, Erickson JW: Structure-based inhibitors of HIV-1 protease. Annu Rev Biochem 62:543, 1993

Wlodawer A, Pavlovsky A, Gustchina A: Crystal structure of human recombinant interleukin-4 at 2.25 Å resolution. FEBS Lett 309:59, 1992

Wlodawer A, Pavlovsky A, Gustchina A: Hematopoietic cytokines: Similarities and differences in the structures, with implications for receptor binding. Protein Sci 2:1373, 1993

Wlodawer A, Swain AL, Gustchina A: Comparison of crystal structures of inhibitor complexes of the human immunodeficiency virus protease. In: Molecular Aspects of Chemotherapy (Shugar D, Rode W, Borowski E, eds) Springer-Verlag, p 173, 1992

Yew N, Strobel M, Vande Woude GF: Mos and the cell cycle: The molecular basis of the transformed phenotype. Curr Opin Genet Dev 3:19, 1993

Yi T, Gilbert DJ, Jenkins NA, Copeland NG, Ihle JN: Assignment of a novel protein tyrosine phosphatase gene (*Hcph*) to mouse chromosome 6. Genomics 14:793, 1992

Yokoyama T, Copeland NG, Jenkins NA, Montgomery CA, Elder FFB, Overbeek PA: Reversal of left-right asymmetry: A situs inversus mutation. Science 260:679, 1993

Young HA, Jenkins NA, Copeland NG, Simek S, Lerman MI, Zbar B, Glenn G, Ortaldo JR, Anderson SK: Localization of a novel natural killer triggering receptor locus to human chromosome 3p23-p21 and mouse chromosome 9. Genomics 16:548, 1993

Zhou R, Daar I, Ferris DK, White G, Paules RS, Vande Woude GF: pp39^{mos} is associated with p34^{cdc2} kinase in c-*mos*^{xe} transformed NIH3T3 cells. Mol Cell Biol 12:3583, 1992

Enrichment Activities

Photo Album

Enrichment Activities

Conference on Gene Regulation and Carcinogenesis

The Ninth Annual Conference on Gene Regulation and Carcinogenesis was held November 3-6 at Big Meadows Lodge, Skyline Drive, Virginia. Attended by the scientific staff of the Basic Research Program and invited guests, this yearly colloquium promotes collaboration among the participants as they exchange information about their research, review the overall direction of research in the Program, and plan new areas of research.

Seminars

The Basic Research Program contributed to the extensive NCI-FCRDC seminar schedule by sponsoring more than 100 seminars during the past year. Included among these was the Special Seminar Series, which brought to the Center such distinguished scientists as Drs. Max Essex, Alan Wolffe, Erling Norrby, Albert Dahlberg, Stephen Hajduk, Lee Hartwell, Bruce Stillman, Jon Beckwith, Max Gottesman, Jack D. Keene, Marilyn Kozak, Ruth Lehmann, Paul Marks, Paul B. Sigler, Michael Chamberlin, Louis Reichardt, Danny Reinberg, Peter Gruss, Peter Goodfellow, David P. Lane, and Douglas Hanahan. A listing of the seminars follows.

Dr. Max Essex
Harvard AIDS Institute
Harvard School of Public Health
Boston, Massachusetts
"New Approaches to Therapy and Vaccination for HIV"
October 6, 1992

Dr. Delong Liu
Mt. Sinai Medical Center
New York, NY
"Signal Transduction, Transforming and Tumorigenic Potential of Human Insulin-Like Growth Factor I Receptor"
October 6, 1992

Dr. John Milner
Pennsylvania State University
University Park, Pennsylvania
"The Impact of Dietary Selenium and Garlic on Cancer Risk"
October 9, 1992

Dr. Tae Gyu Lee
University of Washington
Seattle, Washington
"Regulation of the Interferon-Induced, dsRNA-Activated Protein Kinase by a Cellular 58-kDa Inhibitor During Influenza Virus Infection"
October 19, 1992

Dr. Kristine Vogel
Institute of Neuroscience
University of Oregon
Eugene, Oregon
"Specification of Neurogenic Potential and Neuronal Phenotypes Among Placode- and Neural Crest-Derived Cells"
October 20, 1992

Dr. John Shaughnessy, Jr.
National Cancer Institute
National Institutes of Health
Bethesda, Maryland
"Molecular Characterization of Atypical Genetic Lesions Associated with Deregulated Expression of c-*myc* in Plasmacytoma"
October 22, 1992

Dr. Mark Baer
University of North Carolina
Chapel Hill, North Carolina
"Pyrimidine Dimer Recognition by the Repair Enzyme Yeast DNA Photolyase (Phr1)"
October 27, 1992

Dr. Zhican Qu
Howard Hughes Medical Institute
Johns Hopkins University School of Medicine
Baltimore, Maryland
"Regulation of Tyrosine Phosphorylation of the Nicotinic Acetylcholine Receptor at the Neuromuscular Junction"
November 10, 1992

Dr. Alan Wolffe
National Institute of Child Health and Human Development
National Institutes of Health
Bethesda, Maryland
"Nucleosome Positioning and Modification: Chromatin Structures That Potentiate Transcription"
November 19, 1992

Dr. Michael Jeffers
New York University Medical Center
New York, New York
"The N-*ras* Proto-Oncogene: Identification of Regulatory Elements and Characterization of a Closely Linked Gene"
November 25, 1992

Dr. Roger Reeves
Johns Hopkins University School of Medicine
Baltimore, Maryland
"Manipulation of Mammalian DNA Cloned as Yeast Artificial Chromosomes"
December 2, 1992

Dr. Erling Norrby
Karolinska Institute
Stockholm, Sweden
"Immunization Against HIV-2 and SIV in *Cynomolgus macaques*"
December 18, 1992

Dr. Soochul Park
University of Michigan Medical School
Ann Arbor, Michigan
"Phosphorylation of GAP by the Src Tyrosine Kinase"
January 6, 1993

Dr. Albert Dahlberg
Brown University
Providence, Rhode Island
"Probing the Functions of Ribosomal RNA in Protein Synthesis"
January 7, 1993

Dr. Susan Gottesman
National Cancer Institute
National Institutes of Health
Bethesda, Maryland
"Role of Small RNA in the Regulation of Capsule Synthesis"
January 19, 1993

Dr. Stephen Hajduk
University of Alabama
Birmingham, Alabama
"RNA Editing in Trypanosomes"
January 21, 1993

Dr. Robert Fletterick
University of California
San Francisco, California
"Protein Engineering: Creating Novel Function for Trypsin"
January 21, 1993

Dr. Bart de Vos
Genentech, Inc.
South San Francisco, California
"Human Growth Hormone and Its Receptors: Structural Studies of Receptor Activation"
January 22, 1993

Mr. Steven Dobrowolski
Cleveland Clinic Foundation
Cleveland, Ohio
"Cell Cycle Requirements for $p21^{ras}$ Activity"
January 27, 1993

Dr. Lee Hartwell
University of Washington
Seattle, Washington
"Checkpoints for DNA Damage Repair in the Yeast Cell Cycle"
February 3, 1993

Dr. Xiao-Feng Zheng
Harvard Medical School
Boston, Massachusetts
"Regulation of Cyclin B/$p34^{cdc2}$ Kinase"
February 4, 1993

Dr. Rafael Brüschweiler
Scripps Research Institute
La Jolla, California
"Structure and Dynamics of Biomolecules Assessed by NMR Relaxation"
February 8, 1993

Dr. Bruce Stillman
Cold Spring Harbor Laboratory
Cold Spring Harbor, New York
"Initiation of Chromosomal DNA Replication"
February 18, 1993

Mr. James Sawitzke
University of Oregon
Eugene, Oregon
"RecF Pathway Recombination in the Presence or Absence of the Lambda *orf* Function"
February 19, 1993

Dr. Cliff Tabin
Harvard Medical School
Boston, Massachusetts
"The Molecular Basis for My Inability to Palm a Basketball"
February 19, 1993

Dr. Jon Beckwith
Harvard Medical School
Boston, Massachusetts
"In Vivo Pathways for Disulfide Bond Formation in Proteins"
February 23, 1993

Dr. Robert Lechleider
Harvard Medical School
Beth Israel Hospital
Boston, Massachusetts
"Signalling by a SH2-Containing Phosphatase"
February 25, 1993

Dr. Axel Scheideg
Max-Planck Institut Für Medizinische-Forschung
Heidelberg, Germany
"H-*ras* p21: X-Ray Structure of Complexes with GTP Analogues for Dynamic X-Ray Crystallography"
February 25, 1993

Dr. Morag Park
McGill University
Montreal, Canada
"Activation of the *Met* Oncogene: A General Mechanism for Oncogenic Activation of Receptor Tyrosine Kinases"
February 26, 1993

Dr. Michael Schlunegger
Ciba-Geigy AG
Basel, Switzerland
"Unusual Features Revealed by the Crystal Structure of TGF-2"
March 2, 1993

Dr. Dajun Yang
Michigan State University
East Lansing, Michigan
"Genetic and Molecular Studies on Malignant Transformation of Human Fibroblast Cell Strains by Carcinogen Treatment and/or Oncogene Transfection: Evidence for Multistep Carcinogenesis"
March 2, 1993

Dr. Max Gottesman
Institute of Cancer Research
Columbia University
New York, New York
"Lambda Meets Its Terminator"
March 9, 1993

Dr. Xiao Min Wang
University of Wisconsin
Madison, Wisconsin
"CSF-Like Activity in Mammalian Cells"
March 16, 1993

Dr. Maureen Murphy
University of Pennsylvania School of Medicine
Philadelphia, Pennsylvania
"Identification of Molecular Changes in Meningiomas"
March 18, 1993

Dr. Bruce Jarvis
University of Maryland
College Park, Maryland
"Evolutionary Significance of Secondary Metabolites"
March 19, 1993

Dr. Martine Couturier
University of Brussels
Brussels, Belgium
"CcdB: A Protein Poison of *E. coli* DNA Topoisomerase II and Its Antidote"
March 23, 1993

Dr. Jack D. Keene
Duke University Medical Center
Durham, North Carolina
"Selection from Randomized Libraries Identifies Growth-Factor Messenger RNAs as Targets for Neuron-Specific RNA-Binding Proteins"
March 25, 1993

Dr. Gregory Barsh
Beckman Center for Molecular and Genetic Medicine
Stanford University Medical Center
Stanford, California
"Role of the Mouse *Agouti* Gene in Pattern Formation, Obesity and Embryonic Development"
March 26, 1993

Dr. A.A. Bogdanov
A.N. Belozersky Institute of Physico-Chemical Biology
Moscow State University
Moscow, Russia
"Messenger RNA Pathway Through the Ribosome"
March 30, 1993

Dr. Marilyn Kozak
University of Medicine and Dentistry of New Jersey
Piscataway, New Jersey
"Modulation of Translation by mRNA Structure in Eukaryotes"
April 1, 1993

Mr. James Clark
University of Glasgow
Glasgow, Scotland
"Preliminary Data on the Flow Cytometric Analysis on c-*met* Expression in Breast Cancer"
April 8, 1993

Dr. Sankar Adhya
National Cancer Institute
National Institutes of Health
Bethesda, Maryland
"A New Role of DNA in Transcriptional Regulation"
April 15, 1993

Dr. Ruth Lehmann
Whitehead Institute
Massachusetts Institute of Technology
Howard Hughes Medical Institute
Cambridge, Massachusetts
"Pattern Formation and Germ Line Determination in *Drosophila*"
April 22, 1993

Mr. David Jackson
Brown University
Providence, Rhode Island
"A Molecular Basis for the Control of Hepato-Specific Transcription by Extracellular Signals and Cell Lineage"
April 26, 1993

Dr. Paul Marks
Memorial Sloan-Kettering Cancer Center
New York, New York
"Inducing Transformed Cells to Differentiate: From the Cell Cycle to Clinical Studies"
April 26, 1993

Dr. Junyi Lei
Emory University
Atlanta, Georgia
"Twitchin, a Calmodulin-Dependent Protein Kinase with Pseudosubstrate Domain"
April 27, 1993

Dr. Craig Rosen
Human Genome Sciences
Gaithersburg, Maryland
"New Insights into the NF-κB/Rel Transcription Factor Family"
April 28, 1993

Dr. Paul B. Sigler
Yale University
New Haven, Connecticut
"Stereochemistry of DNA Target Selection by the Steroid/Nuclear Receptors"
April 29, 1993

Dr. James S. Song
University of Vermont School of Medicine
Burlington, Vermont
"DNA Topoisomerase II and Protein Kinase C: Molecular Targets for Cancer Chemotherapy"
May 3, 1993

Dr. Mieczyslaw Chorazy
Maria Sklodowska-Curie Memorial Institute
Gliwice, Poland
"Chromosomal and DNA Damage in Populations from Polluted Areas of Silesia"
May 3, 1993

Dr. Sarah Jacobs
Medical College of Virginia
Richmond, Virginia
"Characterization of a Second Promoter in Human c-*myb* Proto-Oncogenes"
May 4, 1993

Dr. Robert Weisberg
National Institute of Child Health and Human Development
National Institutes of Health
Bethesda, Maryland
"Antitermination and Termination of Transcription in Phage HK022: Differences from the λ Paradigm"
May 5, 1993

Dr. Nancy Ratner
University of Cincinnati College of Medicine
Cincinnati, Ohio
"The Cell Biology of Type I Neurofibromatosis"
May 7, 1993

Dr. Michael Chamberlin
University of California at Berkeley
Berkeley, California
"A New Model of Transcription Elongation and Termination"
May 7, 1993

Dr. Richard Gourse
University of Wisconsin
Madison, Wisconsin
"What Makes Ribosomal RNA Promoters So Strong?"
May 12, 1993

Dr. Louis Reichardt
University of California
Howard Hughes Medical Institute
San Francisco, California
"Adhesive Interactions That Regulate Development of the Retina and Visual Projection"
May 13, 1993

Dr. Lisa Linske-O'Connell
University of Rochester
Rochester, New York
"Stability Effects of Site-Specific Amino Acid Replacements in Yeast Cytochrome C"
May 14, 1993

Dr. Richard Morgan
National Heart, Lung, and Blood Institute
National Institutes of Health
Bethesda, Maryland
"Gene Therapy for AIDS"
May 14, 1993

Dr. Leonora J. Groves
University of California
San Francisco, California
"Purification and Characterization of a Bacterial Adenine Photoproduct Incising Activity"
May 14, 1993

Dr. Jerzy Konopa
Technical University of Gdansk
Gdansk, Poland
"Interstrand DNA Crosslinking by Anthracyclines, Antracenadiones, 1-Nitroacridines, and Related Compounds"
May 17, 1993

Dr. Greg Dressler
National Institute of Child Health and Human Development
National Institutes of Health
Bethesda, Maryland
"Interactions Between Developmental Control Genes and Suppressor Genes in the Kidney"
May 18, 1993

Dr. Sharon Archer
National Institute of Dental Research
National Institutes of Health
Bethesda, Maryland
"The Structure of Profilin as Determined by NMR Spectroscopy"
May 19, 1993

Dr. Yvonne Rosenberg
Henry M. Jackson Foundation for the Advancement of Military Medicine
Rockville, Maryland
"Immunological and Virological Events in Lymphoid Organs During SIV Infection of Macaques"
May 19, 1993

Dr. Michael Cho
University of California
Irvine, California
"Hepatitis A Virus Growth Properties in Cultured Cells and Poliovirus RNA Replication"
May 20, 1993

Dr. John Gearhart
Johns Hopkins University
Baltimore, Maryland
"Transgenesis in the Study of Downs Syndrome"
May 20, 1993

Dr. Danny Reinberg
University of Medicine and Dentistry of New Jersey
Piscataway, New Jersey
"Regulation of Initiation of Transcription by RNA Polymerase II"
May 21, 1993

Dr. Jonathan Stoye
National Institute for Medical Research
London, England
"Using Endogenous Retroviruses to Access the Mouse Genome: *Fv-1* and *hairless*"
May 24, 1993

Dr. Peter Gruss
Max-Planck Institut für Biophysiologische Chemie
Gottingen, Germany
"Pax Genes and Mammalian Development"
May 24, 1993

Dr. Dhruba Chattoraj
National Cancer Institute
National Institutes of Health
Bethesda, Maryland
"Initiator Pairing and Replication Control: Role of Heat-Shock Proteins"
May 25, 1993

Dr. Kimmo Peltonen
Institute of Occupational Health
Helsinki, Finland
"Internal Dose Monitoring of 1,3-Butadiene Exposure"
May 27, 1993

Dr. Baoen Jiang
Medical College of Georgia
Augusta, Georgia
"Astrocytes Modulate Retinal Endothelial Cell Growth, Differentiation, and Extracellular Matrix Expression"
June 1, 1993

Dr. Peter Goodfellow
University of Cambridge
Cambridge, England
"Sex Determination: How to Make a Man"
June 3, 1993

Dr. Jane A. Little
University of Minnesota School of Medicine
Minneapolis, Minnesota
"Upregulation of Fetal Hemoglobin"
June 7, 1993

Dr. John Eppig
Jackson Laboratory
Bar Harbor, Maine
"Cell to Cell Communication Between Mammalian Oocytes and Companion Somatic Cells"
June 7, 1993

Dr. Michael A. Marletta
University of Michigan
Ann Arbor, Michigan
"Nitric Oxide Synthetase"
June 10, 1993

Dr. Warren Hehre
Wave Function, Inc.
Irvine, California
"Computational Strategies in Organic Chemistry"
June 11, 1993

Mr. Wei Wang
University of Virginia Health Sciences Center
Charlottesville, Virginia
"Regulation of Annexin I Activity by Phosphorylation and Proteolysis"
June 14, 1993

Dr. Thomas W. Cline
University of California
Berkeley, California
"How Fruit Flies Choose Their Sex: The Molecular Mechanism of Chromosome Counting and Developmental Commitment"
June 15, 1993

Dr. Richard Yip
California Institute of Technology
Pasadena, California
"Molecular Genetics of a Morphogenetic Locus, *hindsight*, in *Drosophila*"
June 17, 1993

Dr. Achim Leutz
University of Heidelberg
Heidelberg, Germany
"NF-M and Myb: Combinatorial Activators of Myeloid Genes"
June 21, 1993

Dr. Tom North
University of Montana
Division of Biological Science
Missoula, Montana
"Feline Immunodeficiency Virus: A Unique Model for AIDS Chemotherapy"
June 28, 1993

Dr. David P. Lane
University of Dundee
Medical Sciences Institute
Dundee, Scotland
"The p53 Response to DNA Damage"
July 1, 1993

Dr. Jean Cadet
Centre D'Etudes Nucleaires de Grenoble
Grenoble, France
"Oxidative DNA Damage: Mechanisms of Formation and Measurements in Living Systems"
July 6, 1993

Dr. Barbel Meissner
National University of Singapore Hospital
Singapore
"The Termination of rDNA Transcription in *Xenopus* Oocytes"
July 6, 1993

Dr. John Gearhart
Johns Hopkins University
Baltimore, Maryland
"Transgenesis in the Study of Down's Syndrome"
July 8, 1993

Dr. Elizabeth Mansfield
Johns Hopkins University
Baltimore, Maryland
"The Molecular Cloning and Analysis of the *Drosophila* Proliferation Control Gene Hyperplastic Discs"
July 13, 1993

Dr. Philippe Gallay
Centre Hospitalier Universitaire
Vaudois, Switzerland
"Role of the Lipopolysaccharide Binding Protein in Septic Shock"
July 14, 1993

Dr. Yitzhak Zimmer
Weizmann Institute of Science
Rehovot, Israel
"Complexity of FGF Receptors: Genetic Basis for Structural Diversity and Functional Specificity"
July 15, 1993

Dr. Fred Hassner
University of Bar Ilan
Ramet-Gan, Israel
"Stereoselective Dipolar Cycloaddition Leading to Functionalized Rings"
July 16, 1993

Dr. Iva Pichova
Institute of Molecular Genetics
Academy of Science of Czech Republic
Prague, Czech Republic
"Specificity Mapping of HIV-2, HTLV-1, BLV and M-PMV Proteases by HIV-1 Inhibitors"
July 16, 1993

Dr. Nubuo Shimamoto
DNA Research Center
National Institute of Genetics
Mishima, Japan
"Visualization of Single Molecules of RNA Polymerases Sliding Along DNA"
July 20, 1993

Dr. Dan Garrett
National Institute of Diabetes and Digestive and Kidney Diseases
National Institutes of Health
Bethesda, Maryland
"Toward Automated Protein Structure Determination by NMR"
July 21, 1993

Dr. Hiroji Aiba
Nagoya University
Nagoya, Japan
"New Aspect of Catabolite Repression in *E. coli*: Role of CRP Concentration"
July 21, 1993

Dr. Rosemary Walzen
School of Veterinary Medicine
University of California
Davis, California
"Lipoproteins and the Laying Hen: Form Follows Function"
July 23, 1993

Dr. Kirsten Edington
Beatson Institute for Cancer Research
London, England
"Candidate Senescence Genes in Squamous Cell Carcinoma"
July 26, 1993

Dr. Ruth Hall
CSIRO
North Ryde, Australia
"Integrons: Capturing Genes by Site-Specific Recombination"
August 6, 1993

Dr. Jose Luis Millan
La Jolla Cancer Research Foundation
La Jolla, California
"Immortalized Germ Cells Undergo Meiosis *in Vitro*"
August 9, 1993

Ms. Debra Gilbert
Mammalian Genetics Laboratory
ABL-Basic Research Program
NCI-Frederick Cancer Research and Development Center
Frederick, Maryland
"Development and Application of an Interspecific Backcross Linkage Map of the Mouse Genome"
August 13, 1993

Dr. Douglas Hanahan
Hormone Research Institute
University of California
San Francisco, California
"Molecular Characterization of the Stages in Multi-Step Tumorigenesis"
September 3, 1993

Dr. Konrad Trauth
Max-Planck Institut für Immunbiologie
Freiburg, Germany
"Cloning and Expression of the Murine A-*myb* Transcription Factor Gene"
September 7, 1993

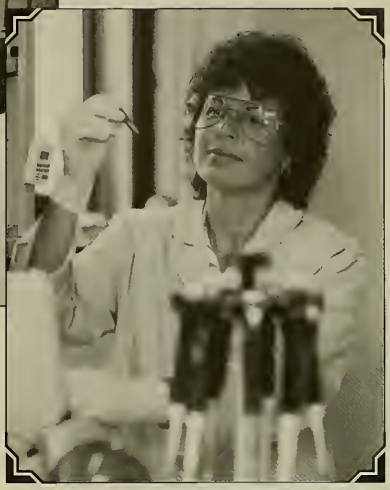
Dr. Karen Steel
MRC Institute of Hearing Research
University of Nottingham
Nottingham, England
"Genetics of Deafness in Mice"
September 7, 1993

Dr. John Cidlowski
University of North Carolina Medical School
Chapel Hill, North Carolina
"Mechanisms of Activation of Apoptosis by Immunosuppressive Drugs"
September 13, 1993

Dr. Klaus Röhm
Philipps-Universitat Marburg
Institut fur Physiologische Chemie
Marburg, Germany
"Towards the Catalytic Mechanism of Bacterial Asparaginases"
September 14, 1993

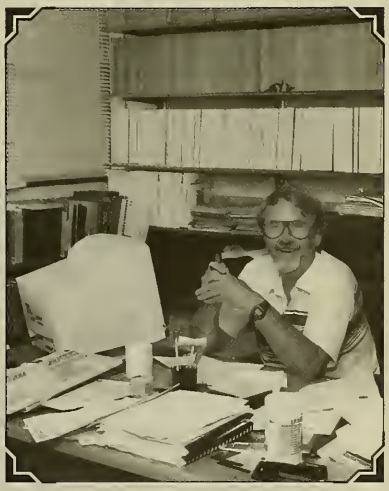
Dr. Vilhelm A. Bohr
National Institute on Aging
National Institutes of Health
Bethesda, Maryland
"Gene Specific DNA Repair: Implications for Mutagenesis, Cancer and Aging"
September 15, 1993

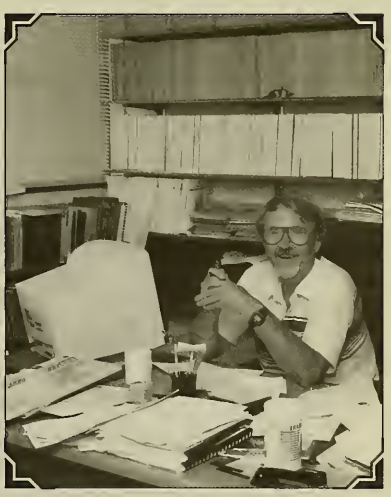
Dr. Linda M. Sargent
McArdle Laboratory for Cancer Research
Madison, Wisconsin
"Specific Chromosome Changes and Alterations in Gene Expression During Progression of Hepatocarcinogenesis"
September 23, 1993



ABL –
BASIC RESEARCH PROGRAM

PHOTO ALBUM 1993





SUMMER STUDENT FELLOWSHIP PROGRAM